THE DEVELOPMENT OF A NEW TECHNIQUE FOR THE NORMALISATION OF THE UNIVERSITY OF CANTERBURY ADAPTIVE SPEECH TEST- FILTERED WORDS (UCAST-FW)

A thesis submitted in partial fulfillment of the requirements for the Degree

Of Master of Audiology

In the University of Canterbury

By J. M. Gibbins

University of Canterbury

2017

Abstract

Low-pass filtered word tests, in which a speech sample is degraded using a low-pass filter (LPF), are one class of low-redundancy test commonly used in the diagnosis of auditory processing disorder (APD). Due to the high level of redundancy within the auditory system and in spoken language, a normal listener is able to fill in the missing speech information and achieve auditory closure even when the speech signal is degraded. The ability to achieve auditory closure is compromised in individuals with APD, which allows filtered speech tests to be used in the diagnostic assessment of APD. One example of this type of test is the University of Canterbury Adaptive Speech Test – Filtered Words (UCAST-FW; O'Beirne, McGaffin and Rickard, 2012). However, the validity and reliability of speech tests are affected by a number of factors, including the homogeneity of the word list. While the UCAST-FW is sensitive enough to discriminate between children with and without APD (Rickard, Heidtke & O'Beirne, 2013), the large variance in the spectral content of its individual test items has resulted in it being somewhat heterogeneous in regards to recognition performance under the same levels of filtering. This creates inherent vulnerabilities within the sensitivity and specificity of a diagnostic test, with increased inter-item variability and reduced inter-patient variability. The present study aimed to compensate for differences in word recognition performance among each word in the UCAST-FW by adjusting the level of filtering such that each word is equally difficult. The performance of 61 English speaking adult listeners with normal hearing was examined on their ability to discriminate speech items both before normalisation (n = 30) and after (n = 31). Psychometric functions (percentage correct vs. LPF frequency) were generated for each word, and were used to calculate relative adjustments for the level of low-pass filtering. These adjustments were performed using a novel method of normalisation that adjusts the levels of low-pass filtering relative to the average performance and takes into consideration the psychometric slope of each of the test words rather than just the midpoint of the function. Results from this study show this normalization technique was successful in achieving a more homogenous word list for both open and closed set testing paradigms, relative to the pre-normalisation testing, as indicated by a more normally distributed cluster of psychometric functions.

Acknowledgments

I would like to thank every one that encouraged me to pursue this path- Libby and Lucy my supporting audiologists who helped me out a lot throughout this whole process.

I would also like to thank my family for encouraging me and supporting me. No matter what I do in life, my mum will always be there to think it's the most amazing thing that anyone in history has ever done, thanks Mum.

To my Creyke Road crew (including the Creyke road rat) it was a blast hanging out with you all, and watching you all struggle through this with me.

Thanks to Greg for doing his job as my supervisor. I really do appreciate all the work you put in to this project. It was a lot of fun working with you.

A special mention to my cousin Brendan who provides endless encouragement by continually asking "How's working at the café going?" despite the fact that I have not worked at a café since 2011, and have completed two postgraduate degrees since then.

Table of Contents

Acknowledgments v

Table of Contents vi

List of Abbreviations xi

Chapter One: 1

1.1 Introduction 1

1.2 Auditory Processing 4
- **1.2.1 Normal central auditory processing** 4
- **1.2.2 Auditory Processing Disorder (APD)** 7
- **1.2.3 Diagnosing APD** 8

1.3 Speech Audiometry 10
- **1.3.1 Word Recognition Speech Tests** 12
- **1.3.2 Low Redundancy Speech Tasks: Auditory Closure and Redundancy** 13
- **1.3.3 Constant vs. Adaptive Stimuli** 15
- **1.3.4 University of Canterbury Adaptive Speech Test –Filtered Words (UCAST-FW)** 15
- **1.3.5 Northwestern University of Children's Hearing in Pictures (NUCHIPS)** 16

1.4 Factors Affecting Word Recognition Performance 18
- **1.4.1 Frequency Content of Speech** 18
- **1.4.2 Neighboorhood Activation Model** 20

1.5 Measures of Word Recognition Performance 22
- **1.5.1 Psychophysical parameters** 22

1.6 Normalisation of Speech Tests 25
- **1.6.1 The Purpose of Normalisation** 25
- **1.6.2 A New Method of Normalisation** 27
  - 1.6.2.1 A rationale: 27
  - 1.6.2.2 Creating a New Method of Normalisation 28

1.7 Statement of the Problem 31
- **1.7.1 Goal of the study** 31
- **1.7.2 Research Questions and Hypothesis** 32

Chapter Two: 34

2 Methods 34
- **2.1 Ethics** 34
- **2.2Recruitment** 34
- **2.3 Participants** 35
- **2.4 Equipment** 36
- **2.5 Stimulus materials** 37
- **2.6 Procedures** 37
  - 2.6.1 Closed set 39
  - 2.6.2 Open set 40
- **2.7 Statistics** 41
  - 2.7.1 Generating Psychometric Functions 41
  - 2.7.2 Statistical Analysis 43

Chapter Three: 45

3 Results 45
- **3.1 Open Set Results** 45

3.1.1 Word Recognition Performance 45
3.1.2 Psychometrically Derived Word Recognition Performance 46
3.1.3 Departures from Normality- SRT 50
3.1.4 Departures from Normality- Slope (%/ octave) 53
3.1.5 Comparative analysis 55
3.1.6 Outliers 61
3.2 Closed set 63
3.2.1 Word recognition performance 63
3.2.2 Psychometrically Derived Word Recognition Performance 64
3.2.3 Departures from Normality- SRT 68
3.2.4 Departures from Normality- Slope (%/octave) 71
3.2.5 Comparative analysis 73
3.2.6 Outliers 78
Chapter Four: 83
4 Discussion 83
4.1 Introduction 83
4.2 Assement of aim one 85
4.2.1 Open set 85
4.2.2 Closed set 87
4.3 Assement of aim two 88
4.3.1 Aim two summary 88
4.3.2 Open Set Outliers 90
4.3.3 Closed Set Outliers 91
4.4 Assement of aim three 91
5 Conclusion 94
6 Limitations and Future Directions 95
7 References 100
8 Appendices 106
Appendix A: Ethical Approval 107
B1: Study advertisment for participant recruitment 108
B2: Information sheet given to each participant prior to testing (page 1 of 3) 109
B3: Consent form signed by all participants 112
Appendix C: Participant Instructions 113
Appendix D: Additional research data 115
D1: Psycometric curves for all words in the UCAST-FW test for pre- and post-normalisation conditions of the open set paradigm. 115
D2: Psycometric curves for all words in the UCAST-FW test for pre- and post-normalisation conditions of the closed set paradigm. 116
D3: Percent deviation from the average performance for SRT and slope values for open set 117
D4: Percent deviation from the average performance for SRT and slope values for closed set 118

List of Figures

Figure 1. Schematic representation of the left (blue) and right (red) primary ascending connections of the human central auditory pathway, with nuclei initials labeled inside boxes. CN; cochlear nuclei; SOC, superior olivary complex; NLL nuclei of the lateral lemniscus; IC, inferior colliculus; MGB, medial geniculate body. The blue representes left side and the red represents the right. Adapted from Schnupp et al. (2011). 5
Figure 2. The influence of extrinsic and intrinsic redundancy on the ability to recognise speech under normal (green) and degraded (red) conditions. 14
Figure 3. A sample response plate from the NUCHIPS picture response book is displayed. The target word is “dog” and the three foils, which are equally familiar to 3-year olds, are “ball”, “car”, and “fr 17
Figure 4. A simplified speech banana with an idealised 500 Hz filter overlying, portraying the audibility of speech sounds under conditions of low-pass filtering. 20
Figure 5. An example of a typical psychometric curve, measuring the proportion of correct responses as a function of SNR (dB). 23
Figure 6. A comparison of three sychometric curves with varying slope functions. Red line shows a steep curve and the blue line shows a shallow curve. The dashed line is a typical psychometric function. 24
Figure 7. A comparison of three psychometric curves with varying slope functions. The red line represents a steep curve and the blue line represents a shallow curve. Word recognition performance at -10 dB SNR for both steep and shallow curves is depicted. 25
Figure 8. The effects of normalizing L_{mid} on similar slope functions (A) and variable slope functions (B). 26
Figure 9. An example of the new normalisation process for the word duck (blue line), against the average (black line) performance, for a 700 Hz filter. As shown on the log-scaled x-axis, the assumed level of filtering is 700 Hz, and the normalised level of filtering for the word duck is 515 Hz, as it generates the same level of word recognition performance as the average at 700 Hz. 29
Figure 10. Display of the UCAST-FW software configuration page. 36
Figure 11. An example display of four-alternative picture choices for the acoustically presented test word “shoe”. Top left: Spoon, top right: school, bottom left: food, bottom right: shoe. 40
Figure 12. Display of the response screen for the open set portion of the UCAST-FW normalisation test. 41
Figure 13. Average word recognition performance across all tested participants and filter frequencies (400, 450, 500, 550, 600 and 800 Hz) for pre-normalisation (blue, *n* =30) and post-normalisation (red, *n* = 31). Word recognition performance for each cut-off filter frequency is averaged across all 50 words in the UCAST-FT word list. Error bars represent standard deviation. 46
Figure 14. Pre- and post- normalisation psychometric curves and post-normalisation curves with significant outliers removed, for the all words in the UCAST-FW open set test. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter. 48
Figure 15. Boxplot of the medians (with interquartile range and outliers) of the distribution of SRT (A) and slope values (B) for open set testing. Distributions of pre- and post normalisation, and post-normalisation with outliers removed are given for both SRT and slope. 52

Figure 16. The percent deviation from the average performance for the SRT and slope of each word in the UCAST-FW. Dashed line represents no deviation from the average. Negative values indicate values greater than the average. A circle represents each word in the UCAST-FW test. Red circles represent values that greatly exceed limits of the x-axis. 58
Figure 17. Average word recognition performance of UCAST-FW for all participants for pre-normalisation and post- normalisation conditions, across the 6 tested levels of low-pass filtering. A circle represents each participant. 59
Figure 18. Psychometric functions for the pre-normalisation outlier food. The blue line represents pre-normalisation performance, the red line represents post-normalisation performance, and the black line represents the average. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter. 61
Figure 19. Psychometric functions for words head, meat, tree and shoe. The blue line represents pre-normalisation performance, the red line represents post-normalisation performance, and he black line represents the average. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter. 62
Figure 20. Average word recognition performance across all tested participants and filter frequencies (200, 300, 350, 400, and 500 Hz) for pre-normalisation (blue, *n* =30) and post-normalisation (red, *n* = 31). Word recognition performance for each cut-off filter frequency is averaged across all 50 words in the UCAST-FT word list. Error bars represent standard deviation. 64
Figure 21. Pre- and post- normalisation psychometric curves and post-normalisation curves with significant outliers removed, for the all words in the UCAST-FW closed set test. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter. 66
Figure 22. Boxplot of the medians (with interquartile range and outliers) of the distribution of SRT and slope values for open set testing. Distributions of pre- and post normalisation, and post-normalisation with outliers removed are given for both SRT and slope. 70
Figure 23. The percent deviation from the average performance for the SRT and slope of each word in the UCAST-FW. The dashed line represents no deviation from the average. Negative values indicate values greater than the average. A circle represents each word in the UCAST-FW test. 75
Figure 24. Average word recognition performance of UCAST-FW for all participants for pre-normalisation and post- normalisation conditions, across the 5 tested levels of low-pass filtering. A circle represents each participant. 76
Figure 25. Psychometric functions for the pre-normalisation outliers door, shoe, and tree. The blue line represents pre-normalisation performance, the red line represents post-normalisation performance, and the black line represents the average. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter. 78
Figure 26. Psychometric functions for words bike, dog, foot, girl, shoe, tree, and watch. The blue line represents pre-normalisation performance, the red line represents post-normalisation performance, and the black line represents the average. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter. 80

List of Tables

Table 1. Pure-tone thresholds (dB HL) for octave frequencies between 250 and 8000 Hz for 61 participants. Values indicate average across both ears. 35
Table 2. The cut-off frequencies for the low-pass filter (LPF) and the number (*n*) of participants who performed the task. 39
Table 3. Psychometrically generated SRT and slope values for each word in the open set UCAST-FW for pre- and post- normalisation testing conditions. Outliers (discussed below) are shown in grey. 49
Table 4. Five highest and five lowest SRT values are given for pre- and post- normalisation conditions for the open set test. 53
Table 5. Five highest and five lowest SRT values are given for pre- and post- normalisation conditions for the open set. 54
Table 6. The coefficient of variation for the word recognition performance among all tested participants for the open set test. 60
Table 7. Psychometrically generated SRT and slope values for each word in closed set UCAST-FW for pre- and post- normalisation testing conditions. Outliers (discussed below) are shown in grey 67
Table 8. Five highest and five lowest SRT values are given for pre- and post- normalisation conditions for the closed set test. 71
Table 9. Five highest and five lowest slope values are given for pre- and post- normalisation conditions for the closed set test. 72
Table 10. The coefficient of variation for the word recognition performance among all tested participants for the closed set test. 77
Table 11. Normalised levels of filtering for all words as a function of the pseudo-levels of filtering (400, 500, 550, 600 and 800 Hz) for open set testing. 81
Table 12. Normalised levels of filtering for all words as a function of the pseudo-levels of filtering (200, 300, 350, 400, and 500 Hz) for closed set testing. 82

List of Abbreviations

UCAST-FW	University of Canterbury Adaptive Speech Test - Filtered Words
APD	Auditory Processing Disorder
FWT	Filtered Words Test
CANS	Central Auditory Nervous System
CAP	Central Auditory Processing
AP	Auditory Processing
MGB	Medial Geniculate Body
LRST	Low Redundancy Speech Test
LPF	Low-Pass Filter
UMAST	University of Canterbury Monosyllabic Adaptive Filtered Speech Test
SRT	Speech Recognition Threshold
SDT	Speech Detection Threshold
NAM	Neighborhood Activation Model
CVC	Consonant- Vowel- Consonant
NU-CHIPS	Northwestern University of Children's Hearing in Pictures
AFC	Alternative Forced Choice
dB	Decibels
dB HL	Decibels Hearing Level
dB SPL	Decibels Sound Pressure Level
IQR	Interquartile Range
SNR	Signal to Noise Ratio
GLM	Generalised Linear Model

Chapter One:

1.1 Introduction

Hearing is an important component of human communication. Auditory processing disorder (APD) is a communication disorder that affects how the brain deciphers and interprets complex sounds. People with APD show symptoms of a peripheral hearing impairment such as experiencing difficulties perceiving complex sounds, particularly in the presence of background noise, while having normal peripheral auditory function. This occurs because there is disruption along the central auditory pathway (Cope, Baguley, & Griffiths, 2015). The disruption can occur at several locations along the auditory pathway, which makes APD a heterogeneous disorder. Children with undiagnosed and untreated APD are often mislabelled as having attention deficit disorders, learning difficulties, being disruptive or inattentive. There is a lack of agreement as to what APD embodies, which makes clinical investigations and interventions difficult. Furthermore, there is currently no gold standard for assessment or intervention, and the testing procedure remains at the discretion of individual clinics. One type of test commonly used to assess APD is the "filtered words test" (FWT), in which a low-redundancy speech sample is distorted by using filtering to modify its frequency content. Low redundancy speech tests assess the ability of an individual to fill in the missing components of a speech signal that is degraded in some way, or presented in the presence of acoustic competition.

Within the auditory system there are multiple neural pathways leading from the VIII nerve to transfer information to higher cortical areas, ensuring that the correct auditory information is passed from the ear to the brain, termed intrinsic redundancy (Whitelaw & Yuskow, 2006).

Additionally, redundancies in acoustic information related to the frequency, intensity and temporal aspects of the speech signal, alongside linguistic knowledge, context and word predictability (Cole & Rudnicky, 1983; Pisoni, 2000) allow a listener to fill in the missing piece of a message, and achieve auditory closure. Degrading or otherwise compromising the acoustic signal reduces extrinsic redundancy; intrinsic redundancy contributes more to the cognitive understanding of the signal. As APD results in a reduction of intrinsic redundancy, auditory closure cannot be achieved when the speech signal is degraded. Thus, the ability of auditory closure is compromised in individuals with APD. This principle has allowed for the production of a plethora of monaural- and binaural-redundancy speech tests to assess central auditory nervous system (CANS) function, including low-pass filtered speech tests. Most available low-pass filtered tests use a constant level of low-pass filtering, such as those fixed at 1000 Hz. This makes them prone to ceiling and floor effects, which reduces the sensitivity and specificity of the test. Research by Keith (2009) found that when the cut-off frequency is set too low, normal listeners would have difficulty, thereby failing to distinguish between normal listeners and those with APD. A recently developed computer-based test, the University of Canterbury Adaptive Speech Test – Filtered Words (UCAST-FW; O'Beirne, McGaffin and Rickard, 2012; Rickard, Heidtke & O'Beirne, 2013), uses an adaptive procedure intended to reduce these ceiling and floor effects, thereby improving the efficiency and sensitivity of the test over its constant level counterparts. However, to further improve the UCAST-FW test and ensure that it is a valid test, a few factors need to be assessed prior to its clinical implementation. When a low-pass filter is applied to speech material, there is potential for the different words to be degraded by different amounts because of the varying spectral content of each individual word. In this sense, some words may become more difficult to understand than other words with the same level of low-pass filtering. This heterogeneity of the test items may affect the reliability and validity of the diagnostic speech test.

The purpose of the proposed study was to ensure that the UCAST-FW word list is homogeneous under conditions of low-pass filtering. The performance of each word was assessed at fixed levels of low-pass filtering. Due to the large variability in psychometric slope parameters between each word in the UCAST-FW list, a novel method of normalisation was created to adaptively change the level of filtering for each word relative to the average performance and the slope of the test word. Within this study, the efficacy of this method to create a more homogeneous UCAST-FW word list under conditions of low-pass filtering was examined.

There is a clear need to normalise the UCAST-FW test to ensure a valid and reliable test prior to clinical implementation, and this study represents an important step in that direction.

1.2 Auditory Processing

1.2.1 Normal central auditory processing

Central auditory processing involves extraction and interpretation of acoustic information from the peripheral auditory system by the central auditory processing system for meaningful recognition of a sound (Griffiths, 2002). The role of the central auditory processing system is to decipher the spectral, spatial and temporal properties of sound (Moore, 2006), through the activity of neurons in many sub and cortical nuclei. This process involves the transformation of sound waves into neural impulses, the transmission of these impulses to higher cortical centers via the auditory nerve, and the cognitive elucidation of these neural impulses into the recognition and perception of the sound (Bamiou, Musiek, & Luxon, 2001; E. M. Elliott, Bhagat, & Lynn, 2007). This neurological processing is a prerequisite for the extraction of meaningful signals from the acoustic environment.

As displayed in Figure 1, auditory processing occurs throughout successive stages of the auditory pathway with hierarchical organisation combining elements of both serial and parallel processing pathways (Rauschecker & Scott, 2009), with each stage having a distinct role in auditory analysis. Each stage within the pathway connects to, and shares information with other nearby and remote nuclei via thousands of neuronal synapses, allowing the auditory system to decode and interpret the input signal (Schnupp, Nelken, & King, 2011). The preceding paragraph provides a general overview of the mechanisms of auditory processing.

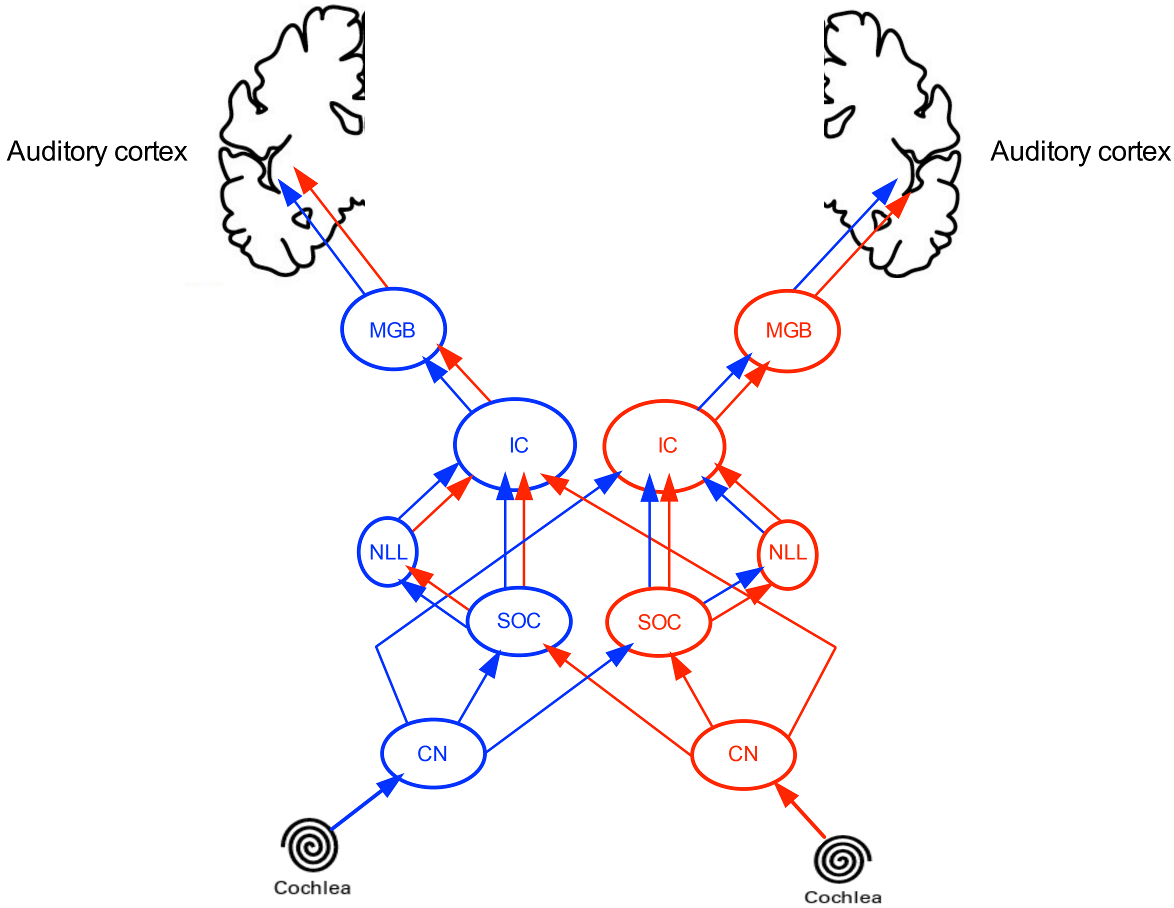


Figure 1. Schematic representation of the left (blue) and right (red) primary ascending connections of the human central auditory pathway, with nuclei initials labeled inside boxes. CN; cochlear nuclei; SOC, superior olivary complex; NLL nuclei of the lateral lemniscus; IC, inferior colliculus; MGB, medial geniculate body. The blue representes left side and the red represents the right. Adapted from Schnupp et al. (2011).

Ascending from the cochlea, temporal and frequency specific information is encoded by the displacement of the basilar membrane, initiating action potentials, which are transmitted to the auditory nerve and to higher centers of the central auditory pathway for subsequent analysis (Fuchs, Glowatzki, & Moser, 2003). The central auditory pathway consists of a neural network of nuclei, beginning at the cochlear nucleus, which decodes and relays information about the timing, intensity and temporal features of the acoustic signal (Caspary, Ling, Turner, & Hughes, 2008). The pathway then extends to the superior olive, which is the first site of binaural convergence. Here, inter-aural timing and level differences between ears are analyzed, which is essential for locating sounds in space (Masterton & Imig, 1984; Tollin & Yin, 2005). Information is then passed along the lateral lemniscus, a tract of axons leading to the inferior colliculus. The inferior colliculus is a mandatory rely station for information from the lower brainstem. As well as these intrinsic projections, the inferior colliculus has an array of connections from the visual and somatosensory centers for multi-sensory integration of sound localisation at the inferior colliculus (Meredith & Stein, 1986). The medial geniculate body (MGB) is the final nuclei for the processing of sensory information before being sent through to the auditory cortex (Bartlett, Stark, Guillery, & Smith, 2000). Due to the large number of projections from several nuclei, the MGB is likely to play a role in sound recognition and localisation, but also the emotional responses to sound (Rees, 2009). At the level of the cortex, the integrated signal links to memory, allowing for association and meaning of sounds to be developed (Weinberger, Ashe, & Edeline, 1994). Combined, processing of auditory information allows for the recognition of auditory objects in relation to the environment, and the assessment of the behavioral significance of the signal.

Along the successive auditory nuclei, signal processing occurs both in a serial and parallel manner, resulting in an efficient and redundant system (Chechik et al., 2006) which also allows for integration with other processes, including attention, memory and language (Cowan, Rosen, Moore, Cacace, & McFarland, 2009; Grossberg, 1999). A lack of information exists between the interactions between auditory processing that occurs from the "top down" (cognitive, language and contextual knowledge) and the processing that occurs from the "bottom up" (extraction of information along the auditory pathway cascade) (Buschman & Miller, 2007) . The interaction between these top-down and bottom-up processes influences how we interpret and use auditory information (Alain, Arnott, & Picton, 2001; Hines, 1999; Samuel, 2001), yet the listening environment determines the contribution of each. In broad terms, auditory processing is the effectiveness and efficiency of the central nervous system in utilising auditory information (American Speech-Language-Hearing, 2005).

1.2.2 Auditory Processing Disorder (APD)

The recognition of and discrimination between complex sounds relies on the extraction and integration of a series of auditory cues that cascade along the auditory pathway (Belin et al., 1998). For some people, the recognition of complex sounds may be difficult, despite the ability of the peripheral auditory system to detect the presence of a sound (Musiek & Chermak, 2013). This is commonly referred to as auditory processing disorder or APD. APD is commonly diagnosed in children, with an estimated prevalence of 5% in the general child population in New Zealand (Esplin & Wright, 2014), but APD can also go unnoticed until later in life. People who have APD will often show pure tone thresholds within the normal range (Cacace & McFarland, 2008) because their peripheral system is functioning normally. However, APD patients often experience difficulties perceiving complex sounds, particularly in the presence of background noise because there is disruption along the central auditory pathway (Cope et al., 2015; Song, Skoe, Banai, & Kraus, 2011). The disruption can occur at several locations along the auditory pathway, which makes APD a heterogeneous disorder (C. A. Miller & Wagstaff, 2011). According to the American Speech-Language-Hearing Association (ASHA, 2005), APD can be due to a weakness in one or many of the following processing abilities: binaural integration, temporal and spectral processing, auditory closure, localisation and lateralisation, and degraded auditory performance in the presence of background noise and/or reverberation. As a result, APD has many clinical manifestations, including difficulties listening in the presence of background noise, difficulties ignoring competing signals, problems recognising and learning new sounds, and difficulties understanding sounds that are distorted or presented in less than optimal conditions (Dawes, Bishop, Sirimanna, & Bamiou, 2008; C. A. Miller & Wagstaff, 2011). Subsequently, children can show symptoms similar to those of a hearing impairment, such as difficulty following oral instructions, being unsure about what they heard, have trouble learning from hearing, and difficulty developing communication skills.

1.2.3 Diagnosing APD

One of the leading issues in contemporary audiology is the underlying etiology of APD (Cacace & McFarland, 1998; Moore, 2006). The problem may arise because of genetic, developmental or idiopathic factors (Palfery & Duff, 2007). Bellis (2002) suggests that there is a relationship between children with chronic otitis media and APD; however, there are many children with severe otitis media with no APD related symptoms. Consequently, there is a lack of agreement as to what APD embodies, which makes clinical investigations and interventions difficult (Cacace & McFarland, 2005). Furthermore, there is currently no gold standard for assessment or intervention, and the testing procedure remains at the discretion of individual clinics. It is also difficult to correctly diagnose APD because of the similarity between the aspects of APD and other language or attention disorders, including autism, attention deficit hyperactivity disorder and other learning disorders (Chermak, Hall, & Musiekl, 1999; Neville, Foley, & Gertner, 2011). The central auditory nervous system plays a role in auditory processing, but also in attention, memory, and language. Children with attention or memory difficulties often have difficulties hearing and following oral instructions. However, in these cases, the auditory neural pathway is intact, and instead it is the child’s attention deficit that impedes their ability to access the auditory information that is presented to them (Bellis, 2004). Therefore, the correct diagnosis of APD related symptoms require a full assessment of the various steps in the auditory pathway. This is accomplished using a test battery in which each sub-test has an emphasis on different aspects of auditory function. Numerous tests have surfaced attempting to correctly diagnose and define APD. Early tests of APD used stimuli such as filtered speech (Bocca, Calearo, & Cassinari, 1954), and dichotic digits (Kimura, 1964), and also included non-speech tests such as sound localisation and temporal-order discrimination. Since then, many commercially available sub-tests have surfaced, designed to assess different aspects of auditory processing. These include the following:

- **Temporal processing tests** which require the ordering or sequencing of successively present acoustic events.
- **Dichotic speech tests**, which evaluate the process of binaural integration or binaural separation.
- **Monaural low-redundancy speech tests,** which evaluate the ability to fill in the missing information of a degraded speech signal.
- **Binaural interaction tests** that assess binaural processes that are dependent on intensity, timing, or differences between two ears.

In 2000, a consensus statement regarding the diagnosis of APD in school-aged children was published in an attempt to reach an agreement on how to diagnose APD (J. Jerger & Musiek, 2000). They constructed a minimal test battery to gather sufficient information necessary for the diagnosis of APD, which consisted of behavioral tests including pure tone audiometry, speech recognition performance, a dichotic digits task, duration pattern sequence testing and temporal gap detection. In addition, electroacoustic and electrophysiological measures, such as immittance testing, otoacoustic emissions, auditory brainstem response, and middle latency response were recommended for additional diagnostic information.

The recommended minimal test battery described above summarizes a reasonable approach to diagnosing APD. However, no definitive protocols have gained wide acceptance. A missing dimension from this test battery includes discriminating between disorders specifically involving the auditory system from other disorders that may influence auditory processing such as ADHD and dyslexia. Other specific problems include the age at which the diagnostic test battery can be administered reliably and the implication for early intervention. There is a clear need for further investigation and research to address this problem.

1.3 Speech Audiometry

An integral part of the diagnostic audiological test battery is speech audiometry. Speech audiometry provides information on an individual's ability to recognise, detect and process speech information, which can be regarded as factors underlying higher order processing and cognitive function. Additionally, speech audiometry also provides supplementary information for appropriate diagnosis and rehabilitation (Meister, 2017). The most commonly used forms of speech audiometry are speech detection threshold (SDT) measures and speech recognition thresholds (SRT). The SDT is the minimum level in dB HL (decibel hearing level) required for a listener to recognise 50% of the speech material, while SRT is the minimum level in dB HL, which allows a listener to recognise 50% of the speech material (American Speech-Language-Hearing, 1988). When developing a speech test, there are several factors to consider for ensuring a reliable and valid test. Hudgins, Hawkins, Kaklin, and Stevens (1947) suggested four guidelines for the development of speech materials for ensuring valid clinical practice:

i. First, speech items should be familiar and simple, thereby removing lexical knowledge or intelligence contributing to the results. Ramkissoon, Proctor, Lansing, and Bilger

(2002) indicated that difficult and unfamiliar speech items result in inconsistent SRT results. Additionally, performance of word recognition has been found to be negatively affected when a talker with a different dialect presents the test items, even when the language is shared (Weisleder & Hodgson, 1989). Weisleder and Hodgson (1989) found that speakers of the native dialect achieved higher scores than did speakers of the same language but of a different dialect.

ii. Second, test items should be phonetically different from one another, avoiding words that rhyme or sound similar. Research by Luce and Pisoni (1998) describe the Neighborhood Activation Model (NAM), the process by which speech items are identified from similar sounding speech items, and how it relates to active memory. The results of their studies conclude that similar sounding words present in a speech test affect the accuracy of results.

iii. Third, Hudgins noted that the word list should contain a normal sampling of speech sounds commonly found in the native language. However, a study by Martin, Champlin, and Perez (2000) have since shown that phonetic balance shows little difference to speech tests with no regard for phonetic balance.

iv. Last, word stimuli need to be homogeneous with regard to audibility. In situations where the speech items are altered such as the application of a low-pass filter, the audibility of each item needs to be equal in order for a valid and reliable speech word list. This can be measured by calculating psychometric threshold and slope, relating to correct recognition as a function of intensity (Nissen et al., 2011).

Thus, there are several factors to consider for developing a reliable and valid diagnostic speech test.

1.3.1 Word Recognition Speech Tests

Word recognition speech tests can either be in an open set format or a closed set format. In an open set test, the listener repeats what was heard without any visual cues or indicators, such as a list of potential options. The listener will be instructed to repeat what they heard, even if what they heard was nonsense syllables. In a closed set test, the number of possible response items is limited, in that the listener will select from a set number of provided alternatives (generally between two and four) to indicate what they heard. This may be verbal, or they point to a visual item. An open test is much more difficult and often results in lower recognition (Madell & Flexer, 2008). A closed set test is often preferred in children as it can be simply and quickly administered and scored (Clopper, Pisoni, & Tierney, 2006) and the results are reliable down to a small number of trials, as demonstrated by consistent test-retest reliability in both a group, and an individual basis (Gelfand, 1998, 2003). The limited number of possible responses available in a closed set test means there is a non-zero chance of choosing the correct response. However, in open set test, performance is related to lexical competition and vocabulary knowledge (Clopper et al., 2006). Thus, the chance performance in a closed set test is fundamentally higher than that of an open set, being equal to 1/n, where n is the number of alternatives. Another fundamental difference between the closed and open set paradigm is the scoring of responses. Single-word tests presented in the closed-set format lend themselves to being scored on a pass/fail basis, often by computer. That is, the respondent either selects the correct visual cue, or they do not. Open set tests are typically scored by the clinicians/researchers themselves, and therefore scoring of consonant-vowel-consonant (CVC) monosyllabic test words is able to be performed on a phoneme-by-phoneme basis. This increases the number of scorable test items by a factor of three. Responses are scored based on the proportion of the word that is correctly identified. This allows for a more precise measure of reception of the acoustical cues of speech and is less affected by non-acoustic factors such as lexical context.

1.3.2 Low Redundancy Speech Tasks: Auditory Closure and Redundancy

The perceptual identification and recognition of speech is an extremely complex set of processes dependent on a large number of factors. These factors can be categorised into those that affect intrinsic redundancy and those that affect extrinsic redundancy (Teatini, 1970). Intrinsic redundancy relates to the multiple neural pathways within the auditory system, and extrinsic redundancy is acoustic information related to temporal, frequency and intensity characteristics of speech, and linguistic knowledge, context and word predictability (Cole & Rudnicky, 1983; Pisoni, 2000). Both intrinsic and extrinsic redundancies ensure that auditory information is passed along to the brain for comprehension of important information, even when the signal is presented in less than optimal conditions (Pisoni, 2000). Low redundancy speech tests (LRSTs) evaluate the ability of an individual to fill in the missing components of a speech signal that is degraded in some way. LRSTs are presented using speech samples that have been altered in order to degrade the signal and make speech perception more difficult. When the acoustic signal is degraded or otherwise compromised, thereby reducing extrinsic redundancy, intrinsic redundancy contributes to the cognitive understanding of the signal. This process is depicted in Figure 2. This allows for auditory closure; the ability to fill in missing portions of the auditory signal, to be achieved when the CANS is functioning well. As APD results in a reduction of intrinsic redundancy due to weakened CANS function, the ability to discriminate low redundancy speech items is impaired, and auditory closure cannot be achieved when the speech item is degraded. This principle has allowed for the production of a plethora of monaural and binaural low redundancy speech tests to assess CANS function.

In a LRST, speech samples may be degraded by having the frequency, spectral or intensity characteristics altered, thereby reducing extrinsic redundancy (Sahli, 2009). These tests are sensitive to dysfunction at the level of the brainstem, as well as the primary auditory cortex, and have been used clinically for many years. Bocca et al. (1954) was the first to use distorted speech testing in order to assess auditory processing ability. Here, they concluded that individuals who had a unilateral cerebral cortex lesion had worse discrimination of filtered speech in the contralateral ear. Additionally, Willeford and Billger (1978) were the first to use low-redundancy speech to assess children with learning disabilities. Further research by Jerger (1960) showed that speech discrimination performance was depressed in patients with temporal lobe lesions when a low-pass filter was applied to speech samples. Since these early investigations, there has been considerable research into the use of LRSTs to assess CANS function, with the further classification of LRSTs into filtered, compressed, expanded, interrupted or reverberated speech signals (Musiek & Baran, 1987). Examples of clinically used LRSTs include a low-pass filtered speech tests, time-compressed speech, and the sentence identification test presented with an ipsilateral competing message (Decker & Nelson, 1981; Farah, Brown, & Keith, 2013).

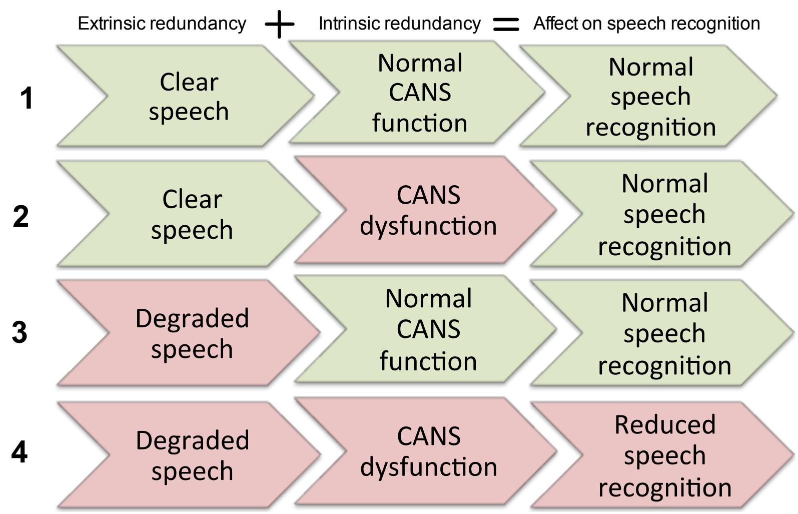


Figure 2. The influence of extrinsic and intrinsic redundancy on the ability to recognise speech under normal (green) and degraded (red) conditions.

1.3.3 Constant vs. Adaptive Stimuli

The current application of low-pass filtered tests are limited in that they use a constant level of low-pass filtering, such as those fixed at 1000 Hz (Rickard, Heidtke, & O'Beirne, 2013). This makes them susceptible to ceiling and floor effects, thereby reducing the sensitivity and specificity of the test. As described above, research by Keith (2009) found that when the cut-off frequency is set to low, listeners with normal hearing and CANS function would have difficulty, thereby failing to distinguish between normal listeners and those with APD. One way to minimise the ceiling and floor effects in a FWT is to use an adaptive low-pass filter, which allows for the degree of redundancy that generates a threshold to be assessed, allowing for a more precise measure of an individual's ability to achieve closure.

One adaptive method is the staircase method (Levitt, 1971), which is similar to the method of limits in that the level of the stimulus is changed depending on the response given by the listener. The value of the stimulus is decreased until the listener provides a negative response. Following which, the value is increased until the listener provides a positive response. As a result, the degree of difficulty of the task is altered, given the listeners previous response. The direction of the runs continues to be increased and decreased until a threshold is reached. An adaptive procedure allows for a more precise test, without encountering ceiling and floor affects. Further, an adaptive approach allows for the degree of redundancy for each individual listener to be assessed, in contrast to a pass/fail paradigm.

1.3.4 University of Canterbury Adaptive Speech Test –Filtered Words (UCAST-FW)

The University of Canterbury Adaptive Speech Test (UCAST) is a software program developed by Dr. Greg O'Beirne using National Instruments LabVIEW 8.20 (O'Beirne, 2009;

O'Beirne, McGaffin, & Rickard, 2012). A previous version of the test was piloted with both children and adults with normal auditory processing skills in order to assess the appropriate parameter configuration as well as the clinical implications of the test for future research. Previous research by O'Beirne et al. (2012) indicated that UCAST-FW test scores were significantly different between adults and children with no known hearing difficulties, indicating a maturation effect. Further research by Rickard et al. (2013) concluded that the UCAST-FW test has the ability to discriminate between children with and without APD with greater sensitivity than its constant-level counterparts. The UCAST-FW test has the potential for clinical applicability, but normalisation of the test is needed to ensure validity and reliability prior to clinical implementation.

1.3.5 Northwestern University of Children's Hearing in Pictures (NUCHIPS)

The Northwestern University of Children's Hearing in Pictures (NUCHIPS) is a closed set picture pointing word recognition test for children, developed by Elliott and Katz (1979). The stimuli from which the NUCHIPS word list was developed consisted of 67 recorded monosyllabic, CVC words and pictures that had been determined to be within the receptive repertories of 3-year olds, and most readily identifiable to 3-year olds when presented at comfortable listening levels. The 67 items were culled from more than 200 words and pictures. Although the NUCHIPS test lists contain 50 words, all 67 pictures are used as foils in the four alternate-forced choice (4-AFC) picture pointing task. An example of a response plate is shown in Figure 3, for the test word "dog". Test foils are randomised within different books so that each set has different arrangements of pictures.

A contributing factor to the recognition of degraded speech is the familiarity of the listener to the words in the test. As APD is ideally diagnosed in younger children, the NUCHIPS word list is designed to be child appropriate. Studies have indicated that children achieve a better word recognition performance when the target items consist of highly familiar words such as house or boat, compared with less familiar word lists such as those in the CID-W, W-22 (Elliott & Katz, 1979; Farrer & Keith, 1981).

The current version of the UCAST-FW utilises the NUCHIPS pictures and word list, with recording of an Australian dialect speaker (National Acoustic Laboratories, Chatswood, NSW, Australia). The current UCAST-FW word list was developed in an American English dialect, and it is likely that the response foils are not transferable to an Australian dialect. Previous research by Murray (2012) was aimed at creating a new set of stimuli for the UCAST-FW that is more appropriate for Australian and New Zealand populations, but this has not yet been incorporated into the test.

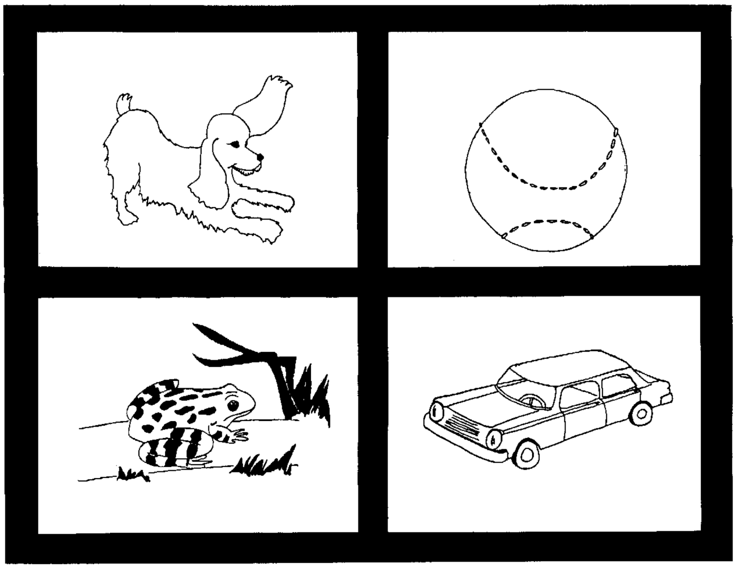


Figure 3. A sample response plate from the NUCHIPS picture response book is displayed. The target word is “dog” and the three foils, which are equally familiar to 3-year olds, are “ball”, “car”, and “frog”.

1.4 Factors Affecting Word Recognition Performance

1.4.1 Frequency Content of Speech

Phonetic contrasts are carried by the temporal and spectral characteristics of speech. Correct phonetic identification requires the use of "phonetic cues" such as duration, formant frequencies, and time-varying amplitude (Winn, Chatterjee, & Idsardi, 2012). By artificially degrading speech samples through removal of spectral content, perceptual phonetic identification can be affected, thus decreasing word recognition ability (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). The normal frequency spectrum of speech ranges from 100 to 8000 Hz, with speech being most intelligible when the bandwidth between 200 and 6000 Hz is available to the listener (Silberer, Bentler, & Wu, 2015). As described above, low redundancy speech tests degrade the speech sample by partially removing specified frequency regions. This is done with the application of spectral filters, attenuating certain frequency components above or below a specified cut-off point. When the low frequencies are removed with a high-pass filter, vowel recognition becomes more adversely affected, for example, making the difference between beet, bat, and bit difficult to interpret (Dorman, Dankowski, McCandless, Parkin, & Smith, 1991). The opposite is true of a high-pass filter, which attenuates the low frequencies. By removing the high-frequency components of speech, it is consonant recognition, which is most adversely affected. For speech understanding, consonant recognition is more important than is vowel (Assmann & Summerfield, 2004). Therefore, the application of a low-pass filter increases the difficulty of speech recognition. However, the attenuation of high-frequency speech cues has varying effect on speech intelligibility, depending on the frequency content of the individual test word. Miller and Nicely (1955) reviewed the effects of low-pass filtering on consonant recognition. They found that the acoustic cues produced by place of articulation are severely affected by a low-pass filter. From these results, Boothroyd (1978) further determined that a small frequency range that contains the most useful acoustic information for consonant identification exists, which correlates to the place of articulation. Therefore, words with the bulk of their acoustic information within the high frequencies will become less intelligible under conditions of low-pass filtering than those words with the bulk of their acoustic information falling within the lower frequencies, with the same level of filtering. A sample of sounds is shown in Figure 4, with an idealized 500 Hz low-pass filter overlaid. This figure shows that low frequency sounds such as /*mm*/ are still audible under filtering, while sounds like /*sh*/ are not. We can reasonably assume that under a 500 Hz filter, words like "meat" will still have a large portion of its acoustic information available to the listener; whereas a word like "shoe" will have less. This means that the same level of filtering will have vastly different effects on word recognition performance depending on the frequency content of the individual word. This is an important factor to consider when using a low-pass filter for diagnostic speech tests, as some words will become significantly harder or easier than other words in the test due to the variation of spectral content and speech cues.

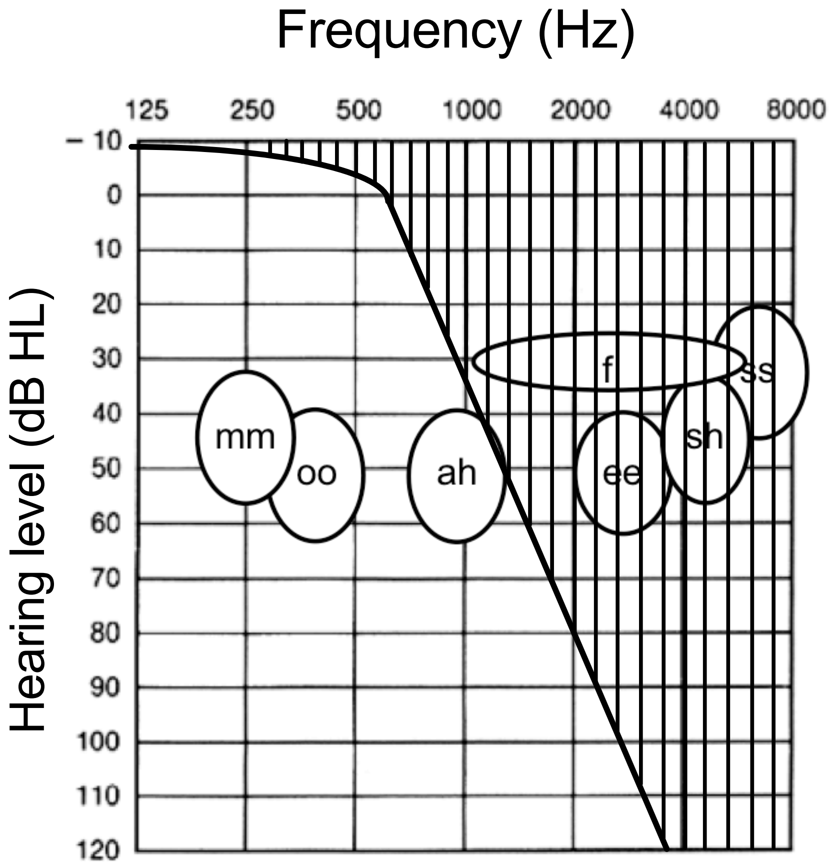


Figure 4. A simplified speech banana with an idealised 500 Hz filter overlying, portraying the audibility of speech sounds under conditions of low-pass filtering.

1.4.2 Neighboorhood Activation Model

Research in both linguistics and psycholinguistics have investigated the role of phonetic information on speech processing. There is a general consensus that word recognition operates on two levels, each resulting in different outcomes for speech processing (Vitevich & Luce, 1998; Vitevitch & Luce, 1999). First, the acoustic-phonetic patterns are processed relative to the similarities of the input signal. This process is explained in the neighborhood activation model (NAM) of auditory word recognition, which describes the process by which a stimulus word is identified in the context of phonetically similar words activated within the memory (Luce & Pisoni, 1998). Research by Vitevitch, Luce, Pisoni, and Auer (1999) on the

NAM showed that density effects, the number of words that are phonologically similar to the test item, can have a negative effect on word recognition performance. Ultimately, word recognition is influenced by the sound patterns of words within the listener's memory and mental lexicon.

Second, research has investigated the effect of probabilistic phonetic information on speech recognition (Storkel & Rogers, 2000). Probabilistic phonetics refers to the frequency that a particular phoneme will occur in a given position of a word (Bailey & Hahn, 2001). Studies have shown that listeners are sensitive to the effects of probabilistic phonetic patterns (Gathercole & Martin, 1996; Treiman, Kessler, Knewasser, Tincoff, & Bowman, 2000). Listeners presented with non-words composed of common phonemes and common sequences of phonemes were rated more likely to be words, than non-words composed of unlikely or uncommon phonemes and sequences of phonemes (Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997).

The combination of these two models can influence word recognition, related to an individual's experience and exposure to particular words. Listeners may be more likely to respond to speech items with words that have higher probabilistic frequency within their mental lexicon. That is, some words may be considered to be more common to some individuals and thus have a stronger mental representation within their lexical neighborhood. Thus, when a speech signal is degraded, phonetic probability and phonetic density overlap to contribute to speech processing and attempt to generate the most likely possibility for word recognition (Bailey & Hahn, 2001; Vitevich & Luce, 1998).

1.5 Measures of Word Recognition Performance

1.5.1 Psychophysical parameters

Measures of word recognition in quiet are most commonly given as an individual's SRT. Because word recognition is a behavioral response, sample data can be used to generate a performance/intensity function, called a psychometric curve. The psychometric curve describes the relationship of an observer's performance to an independent variable, which is generally the quantity of a stimulus in a psychophysical task (Wetherill & Levitt, 1965). Figure 5 shows a typical psychometric function with proportion correct as a function of signal to noise ratio (SNR). To generate a psychometric function, several response thresholds are needed, in which a given stimulus intensity generates performance thresholds. From these data points, the psychometric function can be generated, describing the performance of an observer as a function of some aspect of the stimulus. Within this study, the psychometric function is an '*S*' shaped or sigmoidal curve describing the proportion of word recognition as a function of low-pass filter corner frequency. This was performed for both open and closed set paradigms. As closed set testing is a 4-AFC task with a 25% chance performance, psychometric curves are likely to be more sensitive to chances in stimulus intensity. As described in previous sections, closed set testing is often more reliable and has a higher sensitivity to changes in stimulus intensity as there are limited response options to choose from. This will equate to a more sensitive change in psychometrically derived data with a change in stimulus intensity.

Fitting a psychometric function uses a statistical model using a parameterised family of functions that best match the measured data points. The goal is to fit a model that best predicts the probability of listener performance when the low-pass filter is presented at a specific level.

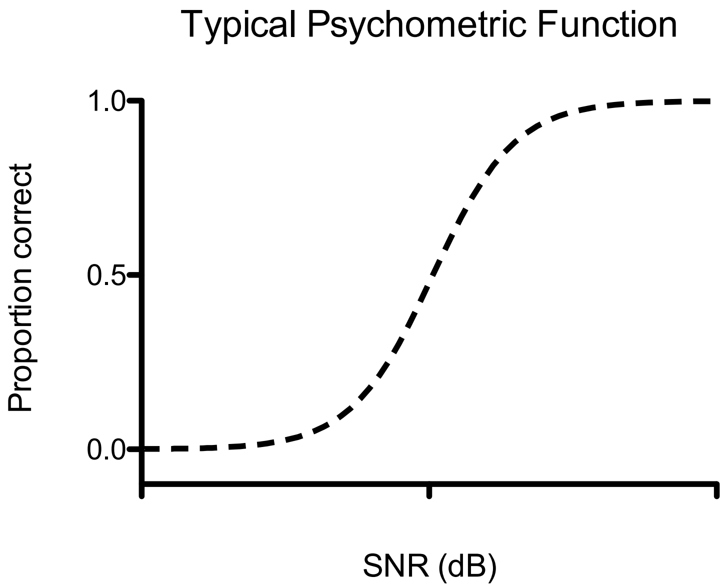


Figure 5. An example of a typical psychometric curve, measuring the proportion of correct responses as a function of SNR (dB).

When developing a psychometric function for word recognition performance as a function of stimulus intensity, there are two primary aspects of the function that needs to be examined (Wilson & Carter, 2001). First is the stimulus intensity that generates a specific word recognition performance, for example, the stimulus intensity that generates 50% word recognition performance. This value may differ depending on the testing paradigm as a SRT is the halfway point between chance performance and 100%. For an open set test, this is still 50%, however, in a 4-AFC closed set test chance performance is 25% which means that a 50% SRT will actually be 62.5% (halfway between chance performance and 100%). The second aspect is the slope of the psychometric curve. The slope of word recognition performance expresses the relationship between the change in correct word recognition performance (Δy) and the change in the presentation level of the signal (Δx), expressed as $\Delta y/\Delta x$ (Wilson & Carter, 2001). Slope functions are often used to determine homogeneity of data, as the slope represents the distributions of word-specific intelligibility, related to the standard deviation. That is, words that generate psychometric curves with a steep slope have smaller distributions of word-specific intelligibilities and thus a smaller standard deviation (Brand & Kollmeier, 2002). Two psychometric curves with different slopes are displayed in Figure 5. The slope is an important factor to consider, as steep slopes and shallow slopes generate very different performance levels with a given stimulus intensity. The greater the difference that exists between slopes, the greater the difference in intelligibility that accompanies a fixed level change in stimulus intensity. In Figure 6, when the stimulus intensity is at -10 dB SNR, the shallower slope has approximately 0.7-proportion correct word recognition, while the steeper slope has complete word recognition. Thus, with the same stimulus intensity, the response rate varies depending on the slope of the psychometric curve.

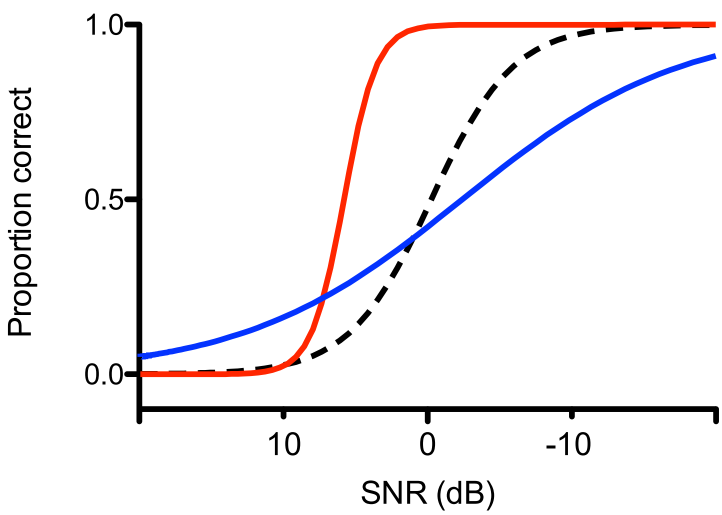


Figure 6. A comparison of three psychometric curves with varying slope functions. Red line shows a steep curve and the blue line shows a shallow curve. The dashed line is a typical psychometric function.

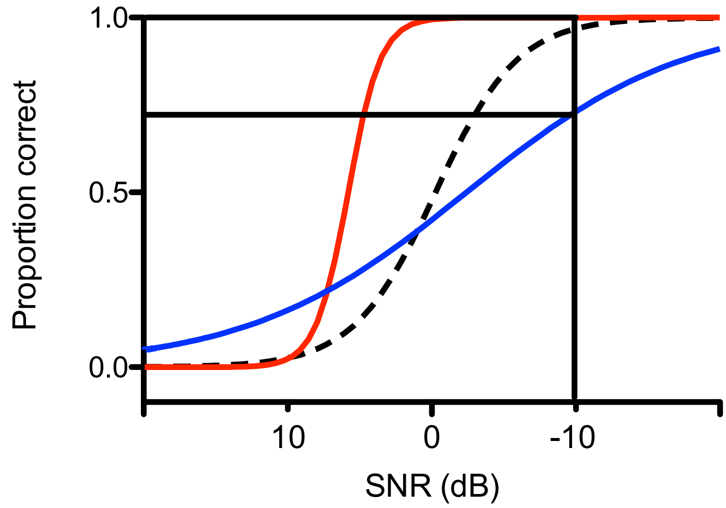


Figure 7. A comparison of three psychometric curves with varying slope functions. The red line represents a steep curve and the blue line represents a shallow curve. Word recognition performance at -10 dB SNR for both steep and shallow curves is depicted.

1.6 Normalisation of Speech Tests

1.6.1 The Purpose of Normalisation

In general, normalisation refers to equalising the difficulty of test materials (Wagener, Josvassen, & Ardenkjær, 2003). Previous methods of normalisation are achieved by manipulating some aspect of the stimulus presentation level, in order for the midpoint (L_{mid}) to overlap. For example, if the pre-normalisation psychometric function for a test item in a speech-in-noise test shows that item to be more difficult than the average, its level may be increased to make it easier and move its SRT closer to the average. Similarly, if the item is easier than the average, its level may be decreased to make it harder and bring it closer to the average. Also, this SRT-adjustment method only creates a homogeneous word list when all slope functions are same, as displayed in Figure 8A. When slope functions are variable, such as in Figure 8B, there are still large variations between slope functions of the test items following normalisation, and a change in stimulus intensity would result in variable word recognition performance depending on the slope, as described above. That is, while the test would be more homogenous when presented near the midpoint of the functions, they become less so when presented at points on the x-axis further from the midpoint. A normalisation processes would, therefore, benefit from a method that takes into account both the slope and the SRT of the psychometric curves, resulting in the same level of word recognition performance with a fixed level increase in the LPF for all words in the test.

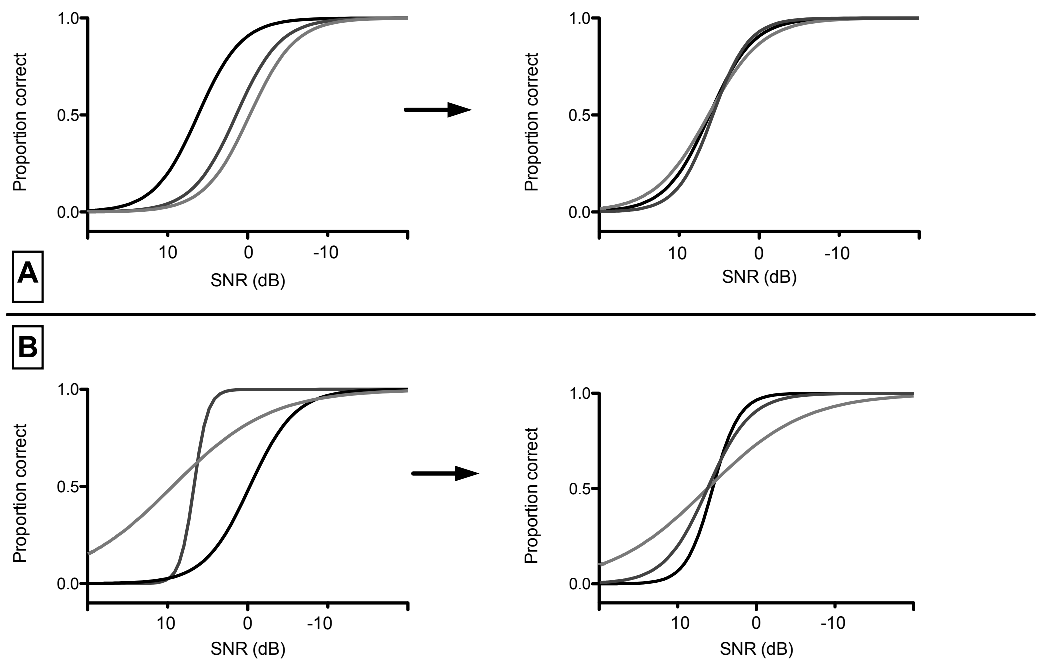


Figure 8. The effects of normalizing L_{mid} on similar slope functions (A) and variable slope functions (B).

It is worth noting that while these level adjustments work for speech-in-quiet and speech-in-noise, it is difficult to apply the same approach to other parameters that affect intelligibility, like low-pass filtering – where the success of pre-filtering or boosting of frequency bands would depend on whether there was high-frequency content to begin with.

1.6.2 A New Method of Normalisation

1.6.2.1 A rationale:

Test items with more uniform difficulty will provide scores that change more readily with changes in stimulus conditions (Ching, Dillon, & Byrne, 1998). The goal of this study is to reduce the variability in psychometric curves for all words in the UCAST-FW, to create a more stable and homogeneous word list under conditions of low-pass filtering. In attempts to create a speech test that is homogeneous in regard to intelligibility, there are two factors of a psychometric curve that need to be examined. First, the cut-off filter frequency that generates 50% correct response, or in the case of the closed set, 62.5% correct word recognition performance. Second is the slope of the function at the midpoint. The slope is an important factor to consider as the UCAST-FW test is administered in an adaptive format, that is, the level of filtering is delivered in a response-dependent manner. This means that variability among slopes will result in different proportions of word recognition performance with a fixed level increase in the low-pass filter for each word. Thus, the greater the difference between slopes, then the change in intelligibility that accompanies a fixed level change in intensity will be greater for words with a steeper slope. Therefore, it is important that a fixed level increase in the low-pass filter will result in the same increase in word recognition performance for each word. By altering the level of filtering based on the normative word recognition performance, we attempt to create a word list that results in similar levels of word recognition performance for all words at a given level of filtering. Thus, creating a truly psychometrically equivalent word list.

A major limitation in using current normalisation methods is the lack of consideration for variation in slope functions. Particularly in low-pass filtered speech tests, there is inherent variability in word recognition performance due to the individual variances in acoustic cues and spectral content of each word. As described above, each word contains vastly different frequency ranges for key acoustic cues, and the application if a low-pass filter can degrade word intelligibility more than other words, depending on the frequency location of these important acoustic cues. Therefore, the normalisation process would benefit from a method that accounts for the variation in word recognition performance among each word, at any given filter frequency.

1.6.2.2 Creating a New Method of Normalisation

Here, we describe a new method of normalisation:

Following measurement of the performance of all 50 words in the pre-normalisation UCAST-FW test (described later), the average psychometric curve was generated. The psychometric curve of any individual test word was compared to this average psychometric curve in order to calculate the adjustment value. The adjustments values are calculated by considering the word recognition performance of the average at a given filter frequency, and determining the filter frequency that the test word needs to achieve the same level of word recognition.

Equation 1 below (adapted from Kollmeier et al., 2015) defines the psychometric function:

$$score = \frac{1}{A}\left(1 + \frac{(A-1)}{1 + \exp\left(\left[-4.\frac{\exp(slope/10)}{100}\right].(\log LPF - \log SRT)\right)}\right) \tag{1}$$

or

Score = (1/*A*)*(1+(*A*-1)/(1+EXP ((-4*(exp.(*slope*/10))/100)*(LOG(*LPF*)-LOG(*SRT*)))))

Where: score = proportion of words correct; LPF = Low-pass filter frequency; SRT = midpoint LPF of function; A = Number of alternatives; slope = slope of the function at the midpoint in %/octave. For the four-alternative closed-set test, A is set at 4, while for the open-set test, A is set to be a very large number (e.g. 100,000 or 1,000,000).

Given a particular proportion correct, we can rearrange Equation 1 to find the LPF that achieves this result. This is shown in Equation 2 below:

$$LPF = 10^{\left(\log SRT + \left[\frac{\ln\left(\left[\frac{A-1}{(score.A)-1}\right]-1\right)}{-4.\frac{\exp(slope/10)}{100}}\right]\right)} \tag{2}$$

Or

LPF = 10^(LOG (*SRT*)+(LN (((*A*-1)/((*score***A*)-1))-1))/(-4*((exp.(*slope*/10))/100)))

Where: score = proportion of words correct; LPF = Low-pass filter frequency; SRT = midpoint LPF of function; A = Number of alternatives; slope = slope of the function at the midpoint in %/octave.

An example is given in Figure 9. In this example, the psychometric curves for the word duck and the average are compared.

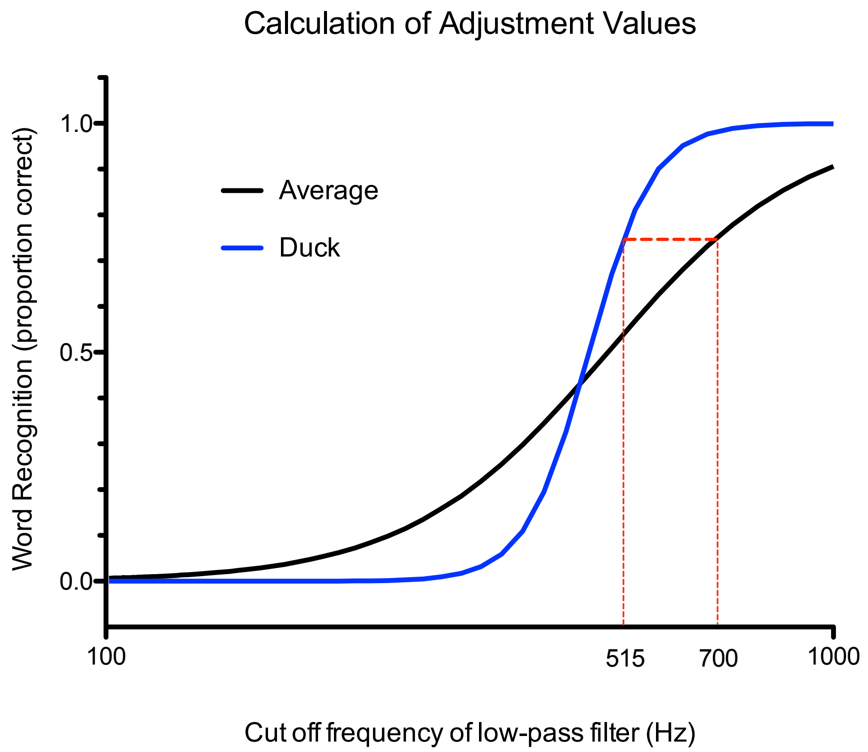


Figure 9. An example of the new normalisation process for the word duck (blue line), against the average (black line) performance, for a 700 Hz filter. As shown on the log-scaled x-axis, the assumed level of filtering is 700 Hz, and the normalised level of filtering for the word duck is 515 Hz, as it generates the same level of word recognition performance as the average at 700 Hz.

The process relies on us knowing the SRT and slope of both the average curve and for the target word “duck”. Here, Equation 1 tells us that if we were to filter the word duck at 700 Hz, we would achieve almost complete word recognition, while the average word would only achieve 0.7 (or 70%). We should aim, therefore, to use a filter frequency for the word duck that will achieve that same 0.7-word recognition performance. To do this, we can then use Equation 2 to calculate the adjusted level of filtering in order for the word duck to achieve the same word recognition performance as the average at 700 Hz, thus creating an adjusted “700 Hz equal” level of filtering for the word duck. As shown in Figure 9, to achieve word recognition performance equal to the average at 700 Hz, we must filter the word duck at 515 Hz.

This process is repeated on the fly for all words in the test, resulting in the same level of word recognition performance when filtered at any given level. The actual value of filtering is dependent on each test word, but will provide an equal word recognition performance for all words in the test. This process allows the level of filtering to be adaptively adjusted depending on the slope of target word relative to the average curve, rather than just a fixed level increase for given LPF frequency.

1.7 Statement of the Problem

1.7.1 Goal of the study

The purpose of this research was to evaluate the homogeneity of the UCAST-FW word list under conditions of low-pass filtering, in order to create a more valid test for the diagnosis of APD. The present study aimed to compensate for differences in word recognition performance among each word in the test by adjusting the level of filtering, in attempts to create a more homogeneous word list in regards to recognition performance as a function of the filter cut-off frequency.

As reviewed above, there are inherent discrepancies in word recognition performance for filtered words. With the use of a low-pass filter, some speech items become increasingly unintelligible relative to others. The large variance among the spectral content of the individual test items creates a word list that becomes heterogeneous in regards to recognition performance under the same levels of filtering. This creates inherent vulnerabilities within the sensitivity and specificity of the diagnostic test. However, no studies have attempted to normalise filtered words speech tests.

A new method of normalisation was created in order to equalise the difficulty of test items by adjusting the level of low-pass filtering in attempts to create a more homogeneous UCAST-FW word list. To the best of our knowledge, this method of equalisation has not been used previously, the approach is individualised for each word, and each tested filter frequency.

1.7.2 Research Questions and Hypothesis

There are three study aims for this project:

Aim 1: *To normalize the difficulty of the UCAST-FW word list by adjusting the level of low-pass filtering, equal to the average pre-normalisation word recognition performance.*

Hypothesis 1: There will be a reduction in the spread of distribution of equivalent SRT and slope values for all words in the UCAST-FW following the normalisation process.

Hypothesis 2: There will be a shift in the SRT and slope values for all words in the UCAST-FW relative to the average pre-normalisation condition following normalisation.

Aim 2*: To determine whether any words in the UCAST-FW are required to be excluded from the UCAST-FW word list.*

Hypothesis 3: There will be outliers present in the post-normalisation conditions for SRT and slope values, indicating poor-performing words that show no improvement following normalisation and will be considered for exclusion from the UCAST-FW word list.

Aim 3*: To determine the differences in word recognition performance in the pre- and post-normalisation conditions between the open and closed set paradigms.*

Hypothesis 4: The closed set results will show a tighter distribution of SRT and slope values, relative to the open set

Hypothesis 5*:* The closed set paradigm will be more sensitive to changes in word recognition performance with changes in the level of low-pass filtering.

Chapter Two:

2 Methods

2.1 Ethics

Ethical approval for this study was obtained from the University of Canterbury Human Ethics Committee, reference 2017/02/ERHEC-LR, as displayed in Appendix A.

2.2 Recruitment

Testing was completed in two groups. Group 1 (n = 30), referred to as the pre-normalisation group, established the normative range of word recognition scores for all the words in the UCAST-FW. Group 2 (n = 31), referred to as the post-normalisation group, performed the same test with a normalised level of low-pass filtering. Sixty-one participants were recruited in total. This number of participants was chosen to obtain an accurate estimation of word recognition performance at various levels of low-pass filtering, as required for calculation of psychometric functions, based on previous experiments involving the UCAST-FW, as well as time and funding constraints. Each participant was given information on the purpose of the study, the amount of time each test will take, and informed that participant information is anonymous. Inclusion criteria for participants were chosen to ensure an accurate representation and normal distribution of the data collected. In order to be included in the study, participants had to be between the ages of 18-40, with air conduction hearing thresholds within normal limits between 250-8000 Hz bilaterally, English was their first language, and the participants had no known learning or medical difficulties that could affect the test, including learning and language difficulties. All participants were offered an inducement of a $10 Motor Trade Association (MTA) voucher as compensation for their time.

2.3 Participants

Sixty-one English-speaking adults participated in this experiment. Group 1 consisted of six males and 24 females participants, with an age range from 20 to 37 years ($M = 26.4 \pm 4.45$). Group 2 consisted of 11 males and 24 females, with an age range from 18 to 36 years ($M = 22.2 \pm 3.69$). There was no attempt to control for gender in this study as previous studies indicate no significant difference in results between male and female listeners in tests for APD (Keith, 2000). Participants were students from the University of Canterbury and volunteers from outside of the university. Hearing thresholds for octave frequencies from 250 to 8000 Hz were within normal limits bilaterally; with the exception of one participant having an 8000 Hz threshold at 35 dB. However, as all test stimuli were low-pass filtered below at least 2 kHz, this 8 kHz hearing impairment was not deemed significant. Summary statistics of the listener's pure-tone thresholds are presented in Table 1.

Table 1. Pure-tone thresholds (dB HL) for octave frequencies between 250 and 8000 Hz for 61 participants. Values indicate average across both ears.

Frequency	*Mean*	*Minimum*	*Maximum*	*Standard deviation*
250 Hz	5.12	-5	25	8.58
500 Hz	3.44	-5	25	6.62
1000 Hz	1.88	-10	20	5.94
2000 Hz	1.76	-10	20	5.51
4000 Hz	0.78	-10	20	5.71
8000 Hz	8.88	-10	35	6.50

2.4 Equipment

For audiological assessment, a GSI-61 (Granson-Stadler Inc.) audiometer was used with TDH-50P super-aural headphones. All audiometric testing was conducted in a sound-treated booth at the University of Canterbury Communication Disorders department, in accordance with the university's audiology protocols and guidelines. The experimental research was conducted in a research laboratory at the University of Canterbury Communications Disorders department.

The UCAST-FW LPF Normalizer v1.01 software (O'Beirne, 2016) was installed on a University of Canterbury personal computer (PC). This software allowed for adjustment of the level of low-pass filtering, presentation level, scoring participant responses for open set testing, and recording and storing results. Software configuration screen is shown in Figure 10**.** The PC had a dual monitor set-up, with the second screen being a touchscreen. The touchscreen was used for the picture pointing exercise in the closed test portion of testing.

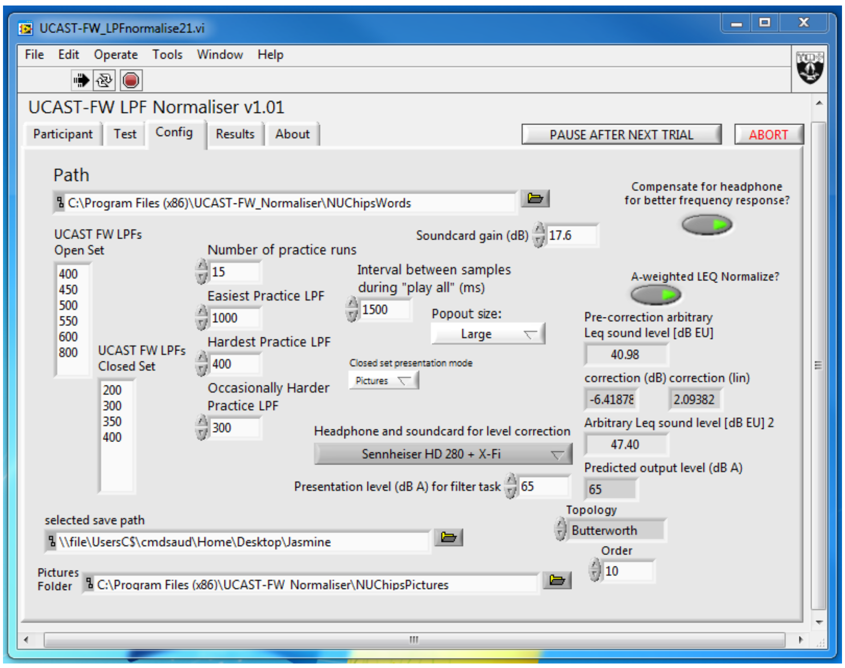


Figure 10. Display of the UCAST-FW software configuration page.

2.5 Stimulus materials

Recording of the Northwestern University Children's Perception of Speech (NUCHIPS) test was taken from the "Speech Recognition Materials" CD 1 (National Acoustics Laboratories, Chatswood, NSW, Australia). All experimental materials were filtered versions of the NUCHIPS test word list, which consists of a 50-word word-list, using an Australian speaker. Elliott and Katz (1979) reported that these words could be used reliably with subjects having receptive vocabulary ages as low as 2.6 years of age. All speech items were delivered and controlled by software developed by Dr. Greg O'Beirne using LabView 8.2. The stimulus was delivered binaurally through Sennheiser HD280 headphones.

All words in the UCAST-FW were subject to low-pass filtering performed using s a 10^{th} order Butterworth filter, designed to pass frequencies below the specified rejection threshold at a rate of 32 dB/octave. Presentation levels were equalized using an LEQ normalisation calculation, resulting in a single dB value that represents the total sound energy over time during stimulus presentation. The target output level was 65 dBA, well above participant hearing thresholds to ensure that audibility did not affect word recognition or discrimination. Ambient noise was less than 40 dB A during the testing procedures.

2.6 Procedures

Both experimental groups were subject to the same procedures.

Custom software was developed to assess word recognition performance at various levels of low-pass filtering. Testing was performed in a sound-treated room at the Communication Disorders Department at the University of Canterbury. Each participant performed both the open set and closed set word recognition test, beginning with the closed set. An example of the instructions given for both sets is displayed in Appendix B. Closed set testing consisted of a four alternate forced choice (4-AFC) picture pointing task, while the open set testing required the listener to repeat the word that that presented acoustically. Participants were familiarised with the 50 words in the test prior to each set. For the closed set task, participants were familiarised with both the word and the corresponding picture. For the open set, the written word was visually presented during familiarisation. Both closed and open set testing begun with 15 practice runs, with the easiest practice low-pass filter frequency at 1000 Hz, which was incrementally decreased to 400 Hz. Following the practice runs, each of the 50 words was presented twice, at each of the cut-off frequencies of the low-pass filter. For open set testing, the cut-off frequencies of the LPF were 400, 450, 500, 550, 600, and 800 Hz. For closed set testing, the cut-off frequencies of the LPF were 200, 300, 350, and 400 Hz. These values were chosen following pilot testing to encompass the likely range of the psychometric functions. Due to the chance performance of closed set testing being 25%, filter frequencies were decreased in the closed set, relative to the open set, to increase the difficulty of the closed set test. An additional cut-off frequency of the LPF at 500 Hz for the closed set was added to seven participants in experimental group 1, and six participants for group 2. The cut-off frequencies of the LPF and the number of participants tested are shown in Table 2.

The second group had adjusted levels of low-pass filtering; meaning that the displayed filter frequency did not represent the true value of the cut-off frequency. The adjusted filter values are described as “equal” values. If the specified cut-off frequency is 400 Hz, then the level of filtering is adjusted using Equations 1 and 2 to create a “400 Hz-equal” level of filtering. What this means is that the test word is being presented at a filter frequency that generates the same level of word recognition performance for the average performance at 400 Hz. Adjusted levels of filtering for the specified levels of filtering for the normalisation process are given in Table 11 for the open set and Table 12 for the closed set.

Table 2. The cut-off frequencies for the low-pass filter (LPF) and the number (*n*) of participants who performed the task.

	Cut-off frequency of the LPF (Hz)	*Pre-normalisation (n)*	*Post-normalisation (n)*		*Cut-off frequency of the LPF (Hz)*	*Pre-normalisation (n)*	*Post-normalisation (n)*
Open Set	400	30	31	**Closed Set**	200	30	31
	450	28	31		300	30	31
	500	30	31		350	30	31
	550	30	31		400	30	31
	600	30	31		500	6	7
	800	30	31				

2.6.1 Closed set

A 4-AFC picture-pointing task was administered to participants at various levels of low-pass filtering. Participants were seated facing a touch screen monitor. Prior to administration of the test, all 50 words and their corresponding images were presented at roughly 1.5 second intervals, in order to familiarise the participant with the words and pictures in the test. Following familiarisation, the testing phase commenced, and the listener was instructed to point to one of the four pictures displayed on the monitor that corresponded to the word that was presented acoustically. An example display of four alternative choices are shown in Figure 11**,** for which the word presented acoustically was “shoe”. Responses were recorded via a touch screen. The presentation order of the 50 words in the word list was randomised, and location of the four pictures alternated randomly. Scores were recorded as either a zero for an incorrect response, or a one, as a correct response.

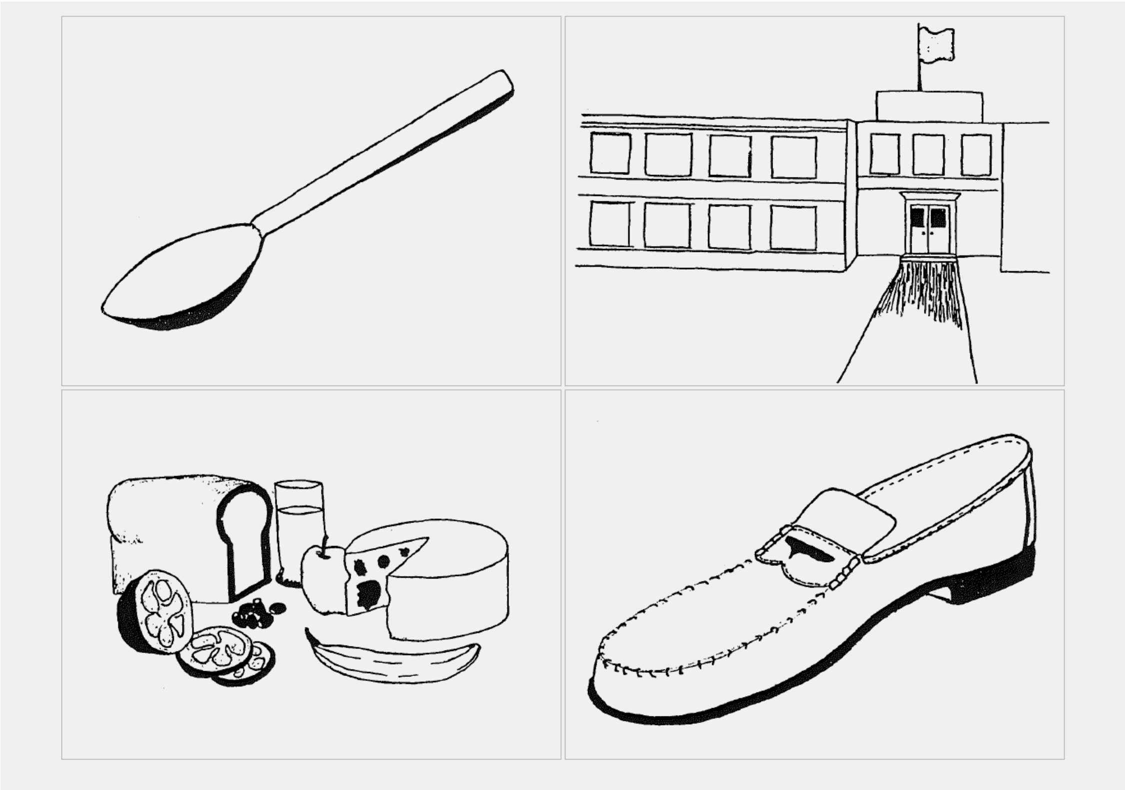


Figure 11. An example display of four-alternative picture choices for the acoustically presented test word “shoe”. Top left: Spoon, top right: school, bottom left: food, bottom right: shoe.

2.6.2 Open set

Open set testing required the listener to repeat what they believed to be the word that was presented to them acoustically, without any visual prompts. Prior to testing, each word in the list was presented binaurally in conjunction with the visual display of the word in written form. Participants were then instructed to repeat what they heard, even if what they heard was nonsense syllables. Participants were asked to face away from the examiner response screen to reduce the chance of viewing the correct answer. Word recognition responses were scored similarly to that of CVC speech audiometry scoring. Responses were scored based on the correct identification of the individual consonants and vowels, giving four possible scorers of 0%, 33%, 66%, or 100% correct recognition of an individual word. An example of the response screen is shown in Figure 12. Displayed is the test word “gum”, which was acoustically presented to the listener, along with the level of low-pass filtering (400 Hz) and the response options (initial c, middle V, final C and all correct).

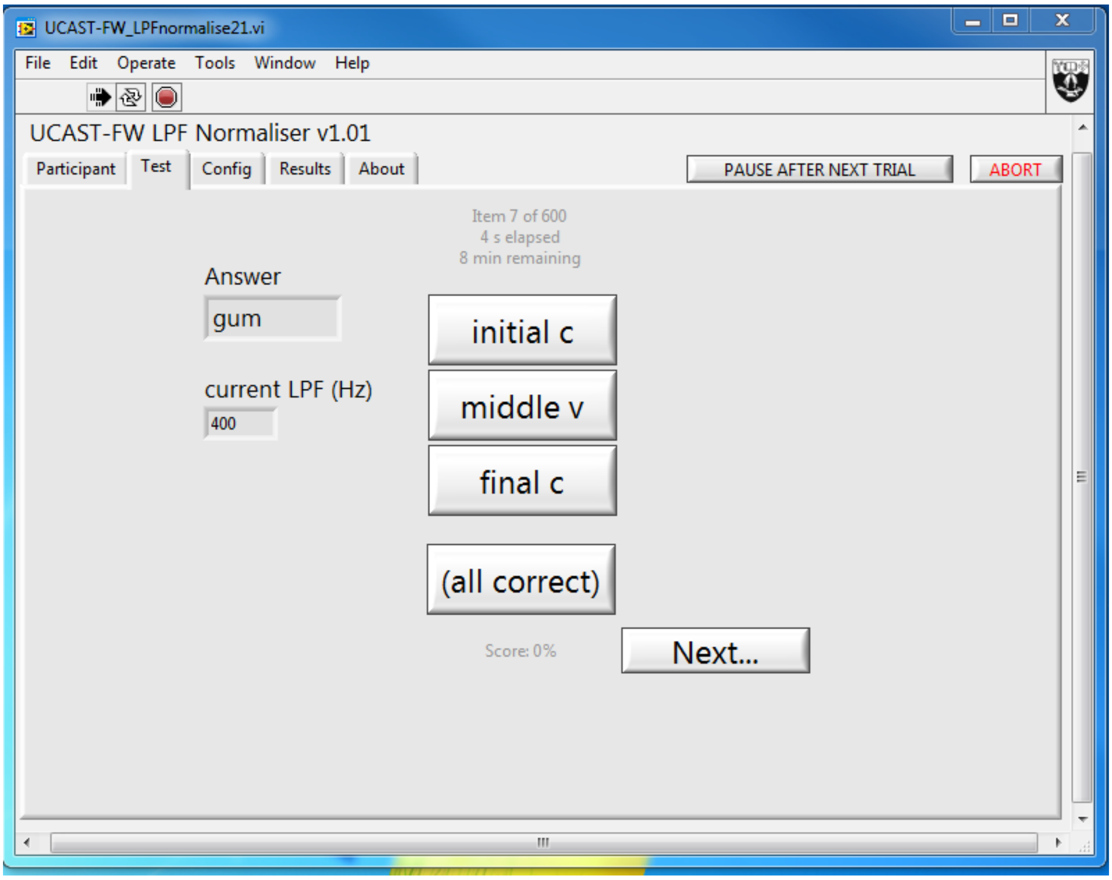


Figure 12. Display of the response screen for the open set portion of the UCAST-FW normalisation test.

2.7 Statistics

2.7.1 Generating Psychometric Functions

Output files produced from the UCAST-FW software following testing were entered into an Excel spreadsheet. Averages, standard deviations, minimum, and maximum values of word recognition performance were calculated.

Psychometric functions are used to summarise the relationship between the performance of word recognition and the cut-off frequency of the low-pass filter. Psychometric functions were generated in R, a computer program for statistical computing. Generalised linear models were fit to collected data points, using the *glm()* function and binomial function in R software. All input values were transformed into log functions, as small changes in the natural log of a behavioral measure are directly interpretable as a percent change. The *glm* function, using a binomial family, provides a powerful and simple method for generating psychometric curves by maximum likelihood. The advantage of using this function is that is can be used to analyse experimental designs with multiple conditions, which is appropriate for this study as multiple levels of low-pass filtering are used. For the binomial family, the response can be specified in three ways (Knoblauch & Maloney, 2012). One of which is specified as a vector of a two-level factor indicating the success/failure of individual trials, such as the success or failure outcomes in the closed set paradigm. Another way is as a numeric vector indicating the proportion of correct responses, such as the responses given in the open set portion of this study. For the open set model, a weight was used to indicate the value of a completely correct response. For the closed set model, the lowest level of word recognition performance is .25, representing chance performance from a 4-AFC task. For this, a special *probit* link function provided in the library '*psych', mafc.probit (4)*, was used. The number '4' indicates that the test used a 4-AFC paradigm. From these models, the cut-off filter frequency that corresponds to halfway between chance and 100% word recognition performance was calculated, which was 0.5 for the open set and 0.625 for the closed set, generating the SRT for each word in the UCAST-FW. Additionally, a value representing the degree of the slope from the generated psychometric function was calculated. These values will be referred to as the slope value from herein, and represents the percent change in word recognition performance as a function of octave frequency (%/ octave).

2.7.2 Statistical Analysis

Quantitative analysis of this study centered on the psychometrically generated SRT and slope values for each word in the UCAST-FW. The qualitative analysis involved visual comparison of the average psychometric function against pre-and post-normalisation conditions, relative to the average psychometric function, for each word in the UCAST-FW.

The distribution of SRT and slope values for the UCAST-FW word list were assessed for normality and outlier estimation, calculated using the "descriptive statistics" function in the IBM Statistical Package for the Social Sciences (SPSS version 24). Potential sources of bias were assessed through skewness and kurtosis, visually inspecting box and whisker plots and conducting Shapiro-Wilk tests of normality, to assess breaches of parametric assumptions. As the data set showed a non-normal distribution of SRT and slope values for both open and closed set, non-parametric testing was performed. All skewness and kurtosis values were transformed into *Z*- scores using the equations below:

$$Z_{skewness} = \frac{S - 0}{SE_{skewness}} \quad Z_{kurtosis} = \frac{S - 0}{SE_{kurtosis}}$$

Note: S is skewness or kurtosis score (generated in SPSS); SE is the standard error.

A *Z*-score comes from subtracting the mean of the distribution (in this case zero) and dividing by the standard error (*SE*) of the distribution (Field, 2013). This was done to create skewness and kurtosis scores that were directly comparable between measures. *Z*-scores for skewness and kurtosis outside ± 1.96 are significant at $p < .05$, and are considered non-normally distributed. A distribution skewed to the right (positive skewness) has a long-right tail with values in the positive direction. The opposite is true for distributions skewed to the left. The distributions of SRT and slope values were graphically depicted using a Tukey box and whisker plot. Differences between pre- and post- normalisation conditions were assessed using a Related Samples Sign Test; this method was preferred, as the distribution of variables did not meet the assumptions for a paired *T*-test or a Wilcoxon Signed Ranked test. A significance level of $p < .05$ was selected and used to evaluate the statistical outcome of the test. Additionally, the aim of the normalisation process was to create a word list that has psychometric functions more analogous to the average, and thus an outcome measure that assessed the degree of difference between a test item and the average was used. The degree of deviation of each SRT and slope values from the average was calculated in Excel, giving a value that represented the percent difference from the average. All values given in Hertz (Hz) are rounded to one decimal place, as hearing is tested in 5 dB steps and small changes in filter frequency down to one decimal place will have no significant impact on word recognition performance.

Chapter Three:

3 Results

3.1 Open Set Results

3.1.1 Word Recognition Performance

The effects of the cut-off frequency of the low-pass filter were assessed as a function of overall word recognition scores for pre- and post- normalisation test conditions. Word recognition performance was scored as the percentage of correct whole words. The percent (%) of correct word recognition for a particular cut-off filter frequency (400, 450, 500, 550, 600 and 800 Hz) was averaged across all 50 words in the UCAST-FW for all tested participants.

The mean word recognition scores for pre-normalisation testing for cut-off filter frequencies of 400, 450, 500, 550, 600, and 800 Hz are 33.9% (*SD* = 19.6), 46.8% (*SD* = 21.5), 56.3% (*SD* = 22.3), 60.4% (*SD* = 22.6), 63.2% (*SD* = 24.3) and 63.5% (*SD* = 24.3), respectively. The mean word recognition score for post- normalisation testing for cut-off filter frequencies of 400, 450, 500, 550, 600, and 800 Hz are 29.9% (*SD* = 13.7), 36.6% (*SD* = 15.1), 45.4% (*SD* = 14.0), 51.5% (*SD* = 12.9), 57.1% (*SD* = 12.5) and 75.2% (*SD* = 16.5), respectively. Results are displayed in Figure 13.

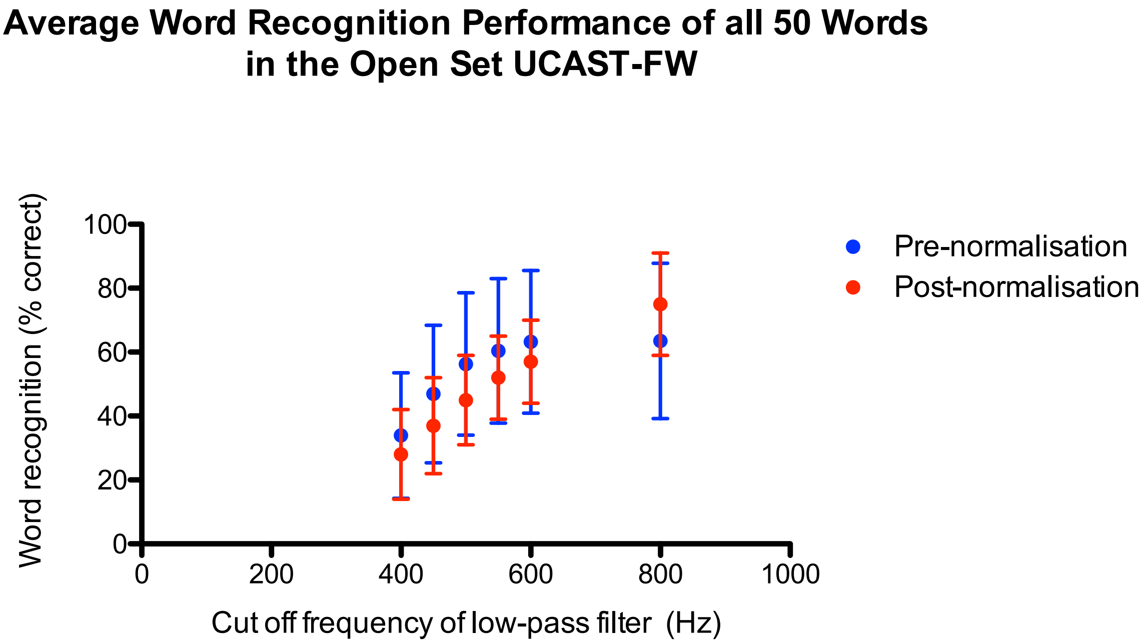
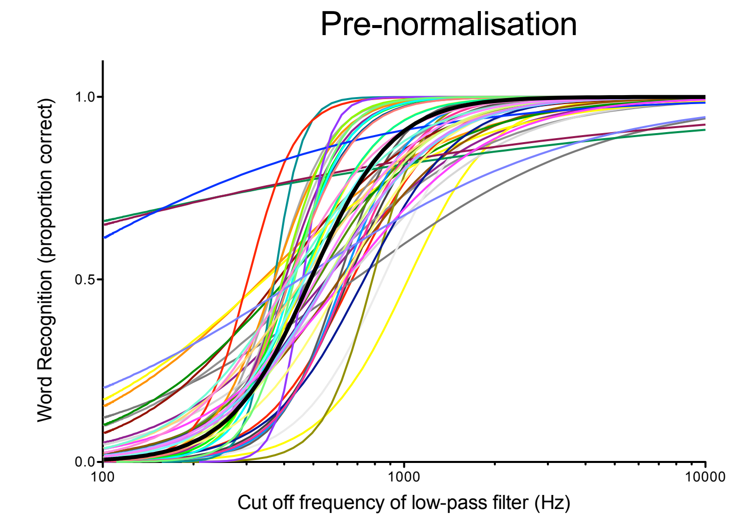


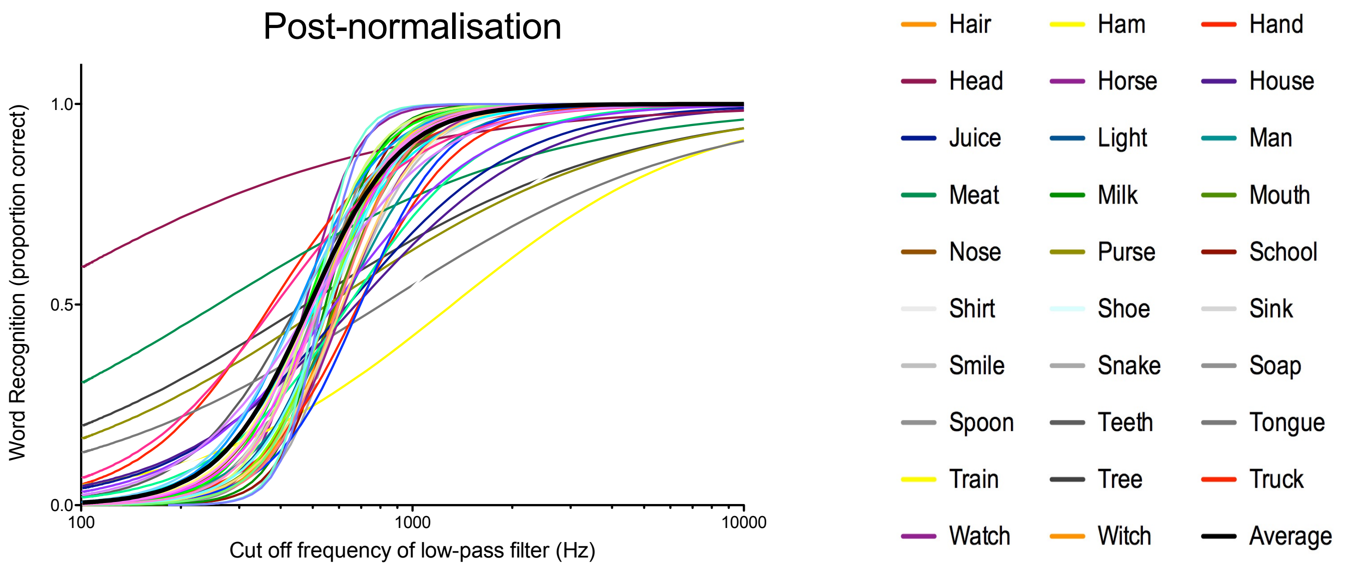
Figure 13. Average word recognition performance across all tested participants and filter frequencies (400, 450, 500, 550, 600 and 800 Hz) for pre-normalisation (blue, *n* =30) and post-normalisation (red, *n* = 31). Word recognition performance for each cut-off filter frequency is averaged across all 50 words in the UCAST-FT word list. Error bars represent standard deviation.

3.1.2 Psychometrically Derived Word Recognition Performance

Following the collection of raw data, logistic regression was used to obtain regression slope and SRT for each word in the UCAST-FW open set paradigm, for both pre- and post-normalisation testing conditions. Figure 14 displays the psychometric curves for all words in the UCAST-FW. Intercept values for each word were calculated to derive the cut-off frequency of the LPF that generated 50% word recognition performance, called the SRT. SRT and slope (%/ octave) values for each word are displayed in Table 3. Overall, the psychometric functions for pre-normalisation conditions generated SRTs that ranged from 15.9 Hz to 1031.3 Hz. (*M* = 490 Hz, *SD* = 184.1). Slope functions ranged from 30.3 to 63.5 (*M* = 50.2, *SD* = 6.7). The psychometric functions for post- normalisation conditions generated SRTs that ranged from 63.2 Hz to 1320.1 Hz (*M* = 556.8 Hz, *SD* = 164.3). Slope functions (%/ octave) ranged from 63.2 to 94.9 (*M* = 85.7, *SD* = 5.6).

Normalisation of the Open Set UCAST-FW Word List





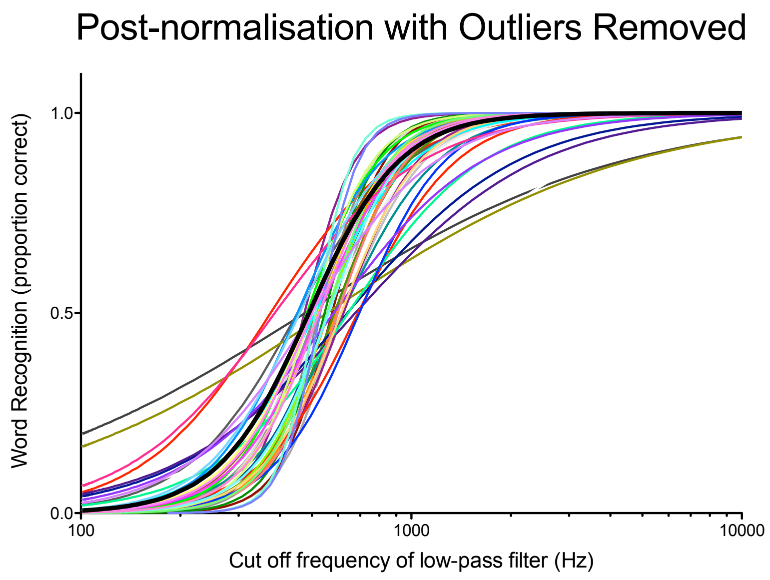


Figure 14. Pre- and post-normalisation psychometric curves and post-normalisation curves with significant outliers removed, for the all words in the UCAST-FW open set test. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter.

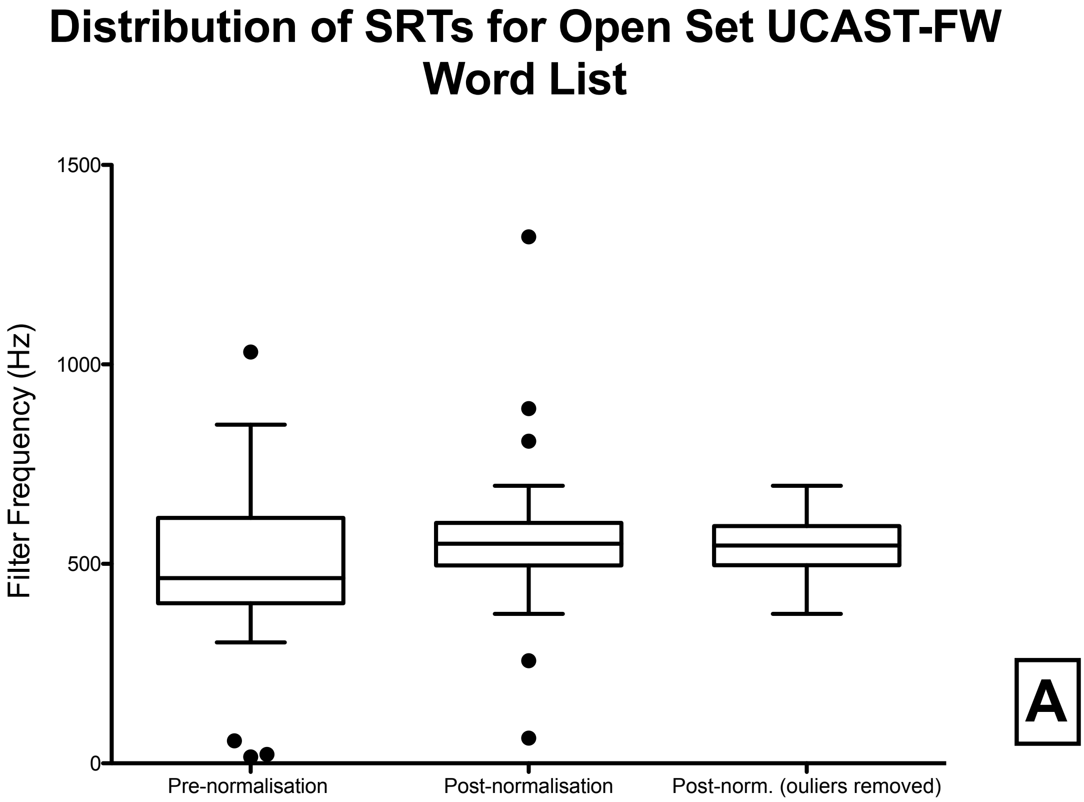
Table 3. Psychometrically generated SRT and slope values for each word in the open set UCAST-FW for pre- and post- normalisation testing conditions. Outliers (discussed below) are shown in grey.

	Pre- normalisation		Post-normalisation	
Word	**SRT**	**Slope**	**SRT**	**Slope**
Ball	411.6	50	525.5	54.1
Bear	509.5	50	500.8	52.9
Bike	560.7	49.4	497.1	48.8
Bird	449.3	39.6	529.2	60.6
Boat	556.8	49.1	459.4	51.8
Bus	467.8	51.5	561.4	54
Cake	420.1	48.8	512.5	61
Clock	433	57.9	556.1	55.8
Coat	553.9	50	494.9	55.8
Comb	627.8	50.1	499.2	52.6
Cup	462.6	48.1	613.7	53.5
Dog	418.4	54.6	598.9	54.9
Door	644.3	53.6	386	47.2
Dress	635.6	46.3	525.5	52.8
Duck	461.3	63.5	581.9	47.1
Food	56.7	38.3	695.9	52.6
Foot	629.6	54.2	452	52.7
Frog	427.5	56.1	511.4	51.2
Girl	458.1	52.9	643.6	48.1
Gum	417.2	55.6	481.8	54.5
Gun	395.4	57.2	557.9	54.8
Hair	466.2	53.1	522.3	54.2
Ham	370.4	55.1	559.8	55.4
Hand	303.4	57.5	689.3	51.2
Head	22.4	31.6	63.2	38.4
Horse	403	55.4	623.2	54.1
House	551.7	49	676.8	45
Juice	737	50.7	640.7	45.8
Light	415.8	54.7	558.1	53.9
Man	371.1	62.8	617	51.5
Meat	15.9	30.3	257.3	39.3
Milk	407.6	45	546.1	57.5
Mouth	517.3	49.3	584.8	54
Nose	784.9	57.3	554.8	40
Purse	610.9	47.8	496.4	53.3
School	381	46.7	575.4	57.5
Shirt	344.5	43.1	889.2	44.4
Shoe	848.8	52	555.2	51.2
Sink	641.3	45.3	490	55.2
Smile	394.9	55.9	622.4	53.2
Snake	372.4	56.1	598.7	57
Soap	482.6	50	514	54.1
Spoon	488.8	43.9	527.5	56.9
Teeth	683.3	40.9	807.8	39.6
Tongue	607.8	55.2	451.9	50.1
Train	654.6	52.9	476.1	39.5
Tree	1031.3	51.7	1320.1	41.9
Truck	672.6	51.4	374.7	48.5
Watch	551.8	45.8	474.3	57.9
Witch	338.6	44	590.8	54.7
Average	489.4	50.2	556.8	51.44
Average (no outliers)	507.5	51.3	550.26	52.63
Min	15.9	30.3	63.2	38.36
Max	1031.3	63.5	1320.1	60.97
Range	1015.4	33.3	1256.8	22.61
SD	184.1	6.7	164.3	5.61
SD (no outliers)	125.4	5.2	80.01	4.46

3.1.3 Departures from Normality- SRT

The distributions of SRT values were tested for normality and homogeneity of variance for both pre- and post- normalisation testing conditions. Pre-normalisation SRT values shows a positive skew to the right with a value of 2.849 and a positive kurtosis value of 2.961. Additionally, a Shapiro-Wilks test was used to test the assumption of normality ($W = 0.938$, $p = .011$). The distribution of pre –normalisation SRTs do not meet the assumption of normality, and show a leptokurtic distribution skewed to the right, indicating that the data set is clustered and has a heavier tail distributed to the right than a normal distribution. Post-normalisation SRT values with outliers included, show a positive skew to the right, with a value of 4.856. SRT values also display a positive kurtosis value of 5.346. Additionally, a Shapiro-Wilks test was used to test the assumption of normality ($W = 0.787$, $p < .05$). The distribution of post-normalisation SRTs with outliers included, do not meet the assumption of normality, and show leptokurtic distribution that is skewed to the right, indicating that the data set is clustered and has a heavier tail distributed to the right than a normal distribution. The removal of outliers will be discussed in the coming sections; a Shapiro-Wilks test was used to examine the assumption of normality for the distribution of SRT values in post-normalisation conditions, with the removal of significant outliers ($W = 0.984$, $p = .797$). The removal of significant outliers following normalisation generated a distribution of SRTs that meet the assumptions of normality, with a skewness of 0.02 and kurtosis of 0.134.

The distribution of SRT values, including outliers, for both pre- and post- normalisation conditions are displayed in a box plot in Figure 15. The highest SRT values indicate those words that require a higher low-pass filter frequency for 50% recognition performance, and thus more speech information, and are therefore commonly more difficult. The opposite is true for words with lower SRTs. The pre-normalisation test words with the highest SRT values are tree, shoe, nose, juice, and teeth. The lowest SRT values, indicating words that are considered easier are meat, head, food, hand, and witch. From this list of extreme values, four were determined to be outliers. An outlier is defines as a value that is either: 1) <25^{th} percentile – 1.5 x interquartile range; or 2) > 75^{th} percentile + 1.5 x interquartile range. Outliers for the open set pre-normalisation conditions are tree (SRT = 1031.3 Hz), food (SRT = 56.7 Hz), head (SRT = 22.4 Hz), and meat (SRT = 15.9 Hz). Post-normalisation words with the highest SRT values are tree, shirt, teeth, food, and hand. Lowest SRT values, indicating words that are considered easier are head, meat, truck, door, and tongue. From this list of extreme values, five were determined to be outliers: tree (SRT = 1320.1 Hz), shirt (SRT = 889.2 Hz), teeth (SRT = 807.8 Hz), meat (SRT = 257.3 Hz), and head (SRT = 63.2 Hz). A list of the extreme values is given in Table 4.



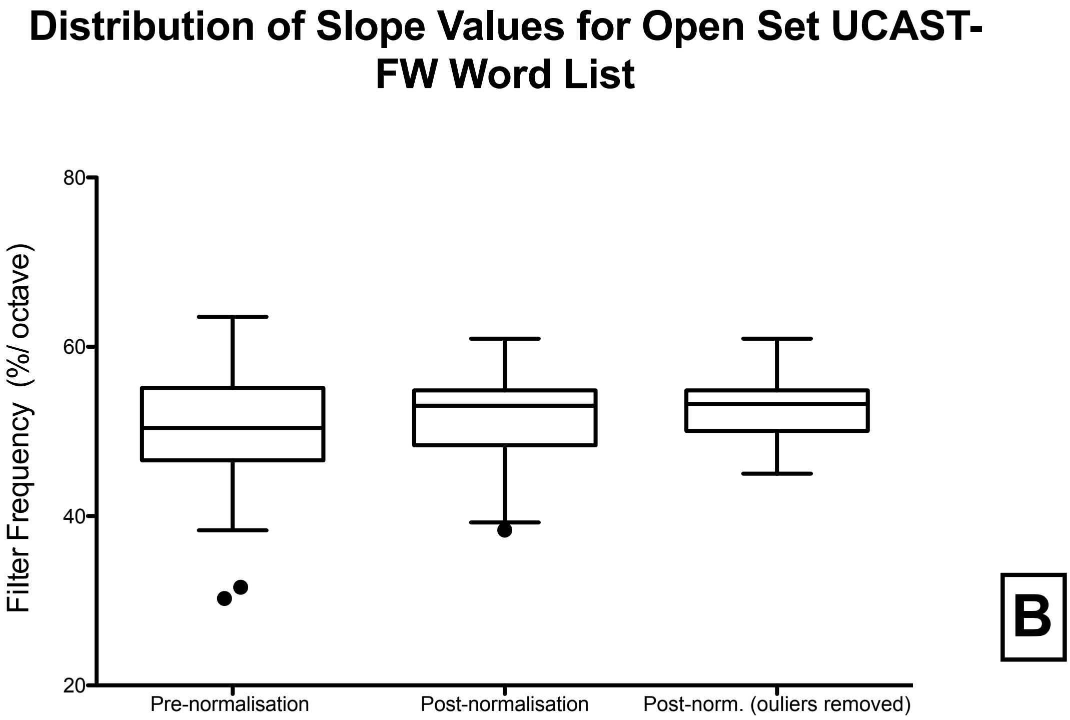


Figure 15. Boxplot of the medians (with interquartile range and outliers) of the distribution of SRT (A) and slope values (B) for open set testing. Distributions of pre- and post normalisation, and post-normalisation with outliers removed are given for both SRT and slope.

Table 4. Five highest and five lowest SRT values are given for pre- and post- normalisation conditions for the open set test. Outliers (discussed below) are shown in grey.

Extreme SRT Values- Open Set						
Pre-norm.		**Word**	**LPF (Hz)**	Post-norm.	**Word**	**LPF (Hz)**
Highest	1	Tree	1031.3	1	Tree	1020.1
	2	Shoe	848.8	2	Shirt	889.2
	3	Nose	784.9	3	Teeth	807.8
	4	Juice	737	4	Food	695.9
	5	Teeth	683.3	5	Hand	689.3
Lowest	1	Meat	15.9	1	Head	63.2
	2	Head	22.4	2	Meat	257.3
	3	Food	56.7	3	Truck	374.7
	4	Hand	303.4	4	Door	386
	5	Witch	338.6	5	Tongue	451.9

3.1.4 Departures from Normality- Slope (%/ octave)

The distributions of slope values were tested for normality and homogeneity of variance for both pre- and post- normalisation testing conditions. Pre- normalisation slope values show a negative skew to the left, with a value of -2.57, and a positive kurtosis value 2.24. Additionally, a Shapiro-Wilks test was used to test the assumption of normality ($W = 0.948$, $p = .028$). The distribution of pre –normalisation slope values do not meet the assumption of normality, and show a distribution with a skew to the left, indicating that the data set is clustered and has a slightly heavier tail distributed to the left than a normal distribution. Post-normalisation slope values show a skewness value of -2.71 and kurtosis value of 0.34, indicating the post-normalisation slope values are negatively skewed to the left. Additionally, a Shapiro-Wilks test was used to test the assumption of normality ($W = 0.904$, $p = .001$). This indicates that both pre- and post normalisation slope values are non-normally distributed. However, a tighter, less spread distribution of slope values are seen in the post-normalisation condition.

The distribution of slope values for pre- and post normalisation conditions are displayed in a box plot in Figure 15. Highest slope values indicate words that have a steeper slope. Conversely, words with lower slope values indicate words that have a shallower slope. Pre-normalisation test words with the steepest slopes are duck, man, clock, hand, and nose. Additionally, the words with the shallowest slope are meat, head, food, bird, and teeth. From this list of extreme values, two were identified to be outliers, which are head (slope = 31.6 %/octave) and meat (slope = 30.3 %/octave). Post-normalisation words with the steepest slopes are cake, bird, watch, school, and milk. Conversely, words with the shallowest slope are head, meat, train, teeth, and nose. From this list of extreme values, one was identified to be an outlier, which is head (slope = 38.4 %/octave). A list of extreme values is given in Table 5.

Table 5. Five highest and five lowest SRT values are given for pre- and post- normalisation conditions for the open set. Outliers (discussed below) are shown in grey.

Extreme Slope Values- Open Set						
Pre-norm	**Word**		**%/Octave**	**Pre-norm**	**Word**	**%/ Octave**
Highest	1	Duck	63.5	1	Cake	61
	2	Man	62.8	2	Bird	60.6
	3	Clock	57.8	3	Watch	57.9
	4	Hand	57.5	4	School	57.5
	5	Nose	57.3	5	Milk	57.5
Lowest	1	Meat	30.3	1	Head	38.4
	2	Head	31.6	2	Meat	39.2
	3	Food	38.3	3	Train	39.5
	4	Bird	39.6	4	Teeth	39.6
	5	Teeth	40.9	5	Nose	40

3.1.5 Comparative analysis

A Related Samples Sign Test indicated a significant difference between pre- and post-normalisation conditions for the open set SRT (Z = -2.121, p = .034). However, no significant difference was seen between pre- and post- normalisation conditions for slope values (Z= -0.424 p = .671). Figure 15 compares the median (with interquartile ranges) and range of SRT and slope values for pre- and post- normalisation conditions. On observation, there is a significant reduction in the spread of SRT values in the post-normalisation condition relative to the pre-normalisation condition. Additionally, a smaller interquartile range between pre-normalisation (IQR = 215), and post-normalisation (IQR = 106.6) is observed, indicating a reduction in the variation of SRT values following normalisation. This trend is also seen in slope values. The distribution of slope values shows a reduction in the spread of values following normalisation, as shown by a reduction in the IQR between pre-normalisation (IQR = 8.6) and post-normalisation (IQR = 6.5).

In order to estimate the deviation of the psychometric functions for each word from the average function, the difference in SRT and slope values from the average for a given word was calculated, and given as a percent difference, displayed in Figure 16. In general, post-normalisation SRT and slope values show a smaller deviation from the average than pre-normalisation values.

The normalisation process can have four effects on word recognition, described below:

1) Improvement: Post-normalisation values become closer to the average performance
2) Deterioration: Post-normalisation values become further from the average than the pre-normalisation performance.

3) No change: Pre-and post- normalisation conditions show a similar percent of deviation from the average; this includes changes from negative to positive values and vice versa i.e. -30% to 30%
4) No effect: extreme deviation from the average in pre-normalistion condition is not resolved

Qualitative analysis shows an improvement in the spread of SRT values, with 29 of the 50 words moving closer to the average performance, relative to pre-normalisation testing. Additionally, six words (bus, cup, duck, girl, house and mouth) show a further deviation from the average in the post-normalisation condition. No change was observed in eleven words (bear, bird, clock, dog, door, hair, horse, soap, spoon, teeth and truck) between pre –and post-normalisation conditions. Of these eleven words, seven (bird, clock, dog, door, hair, horse and truck) show a change from a negative deviation, to a positive; however, the overall percent deviation remains the same. For example, the word truck showed -27% deviation from the average in pre-normalisation conditions, and a 30% deviation in the post-normalisation. Of the 11 words that had no change between testing conditions, three words (bear, soap and spoon) showed SRT values that were within ± 10% deviation from the average for both conditions, indicating the normalisation process would not have changed the levels of low-pass filtering significantly, as these words were already similar to the average performance. A table of values for the percent of deviation for each word is in Appendix D.
Qualitative analysis shows an improvement in the slope values in 19 of the 50 words closer to the average performance, relative to pre-normalisation testing. Additionally 17 words (ball, bear, cake, coat, comb, cup, hair, house, juice, mouth, nose, school, shirt, soap, train, tree, and watch) show a further deviation from the average in the post-normalisation condition. No change was observed in 11 words (bike, boat, bus, dog, door, girl, ham, milk purse, snake and

truck) between pre –and post- normalisation conditions, and three words displayed a large deviation from the average in the pre- and post- normalisation conditions (head, meat, and teeth).

Percent Difference of all Words in the Open Set UCAST-FW from the Average

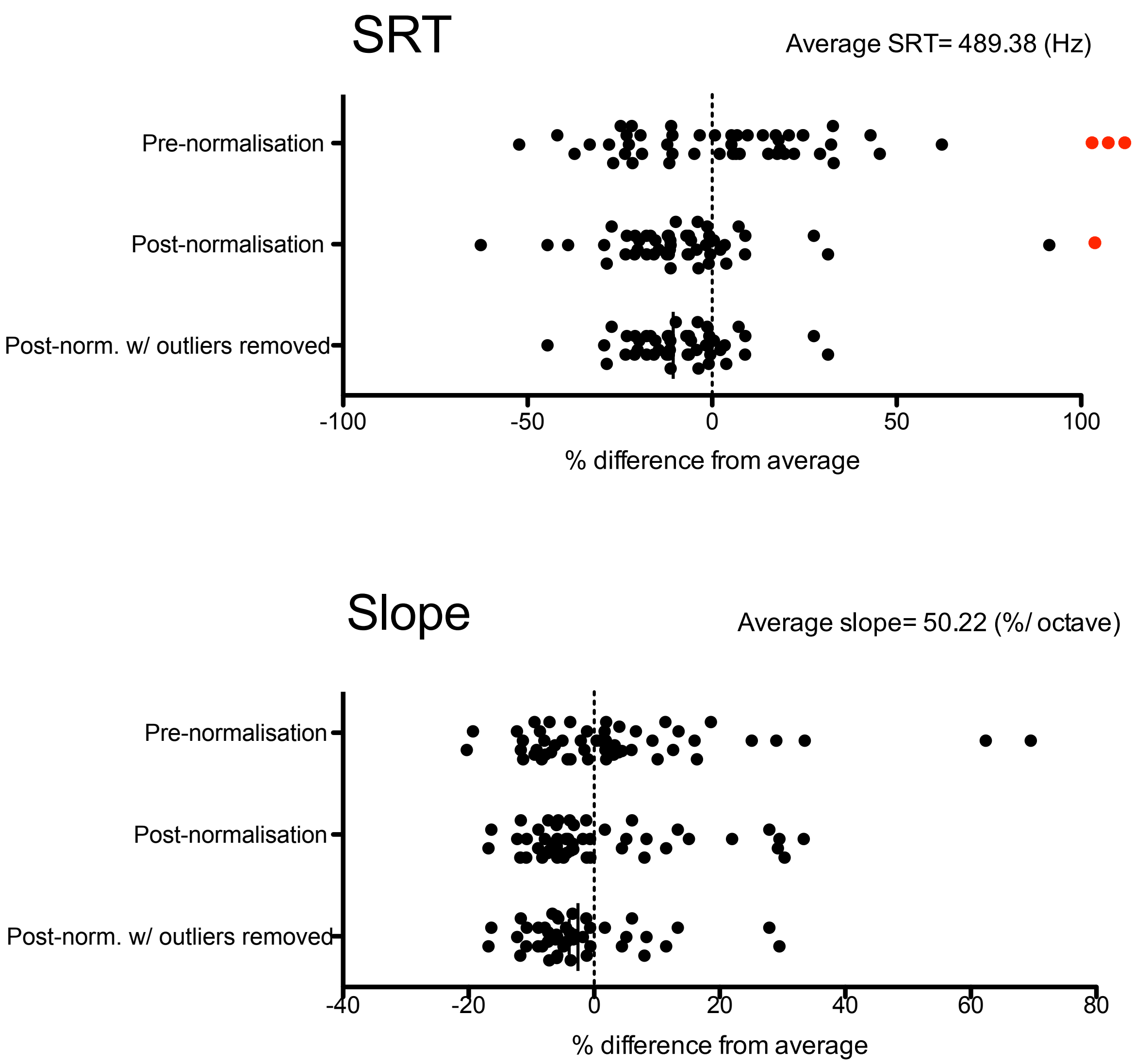
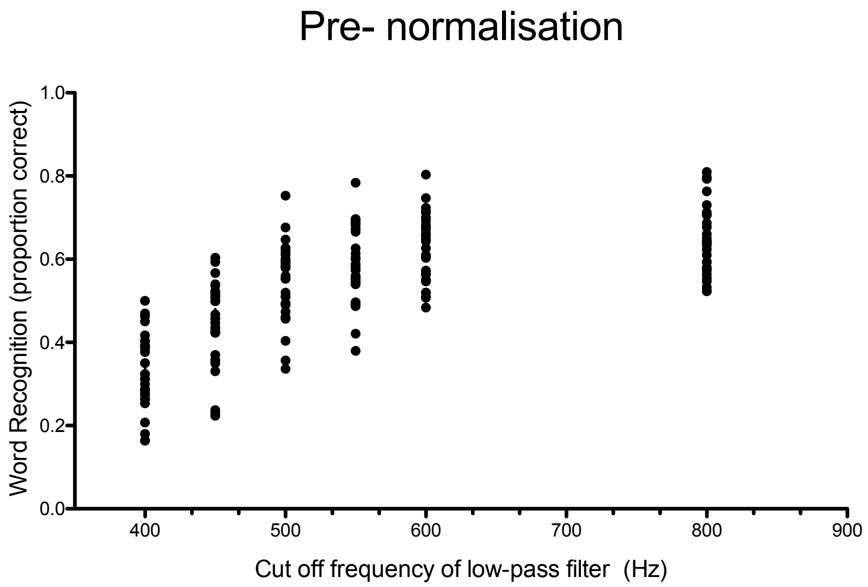


Figure 16. The percent deviation from the average performance for the SRT and slope of each word in the UCAST-FW. Dashed line represents no deviation from the average. Negative values indicate values greater than the average. A circle represents each word in the UCAST-FW test. Red circles represent values that greatly exceed limits of the x-axis.

Effects of Normalisation on Average Intelligibility Scores Across Each Participant for Open Set Testing



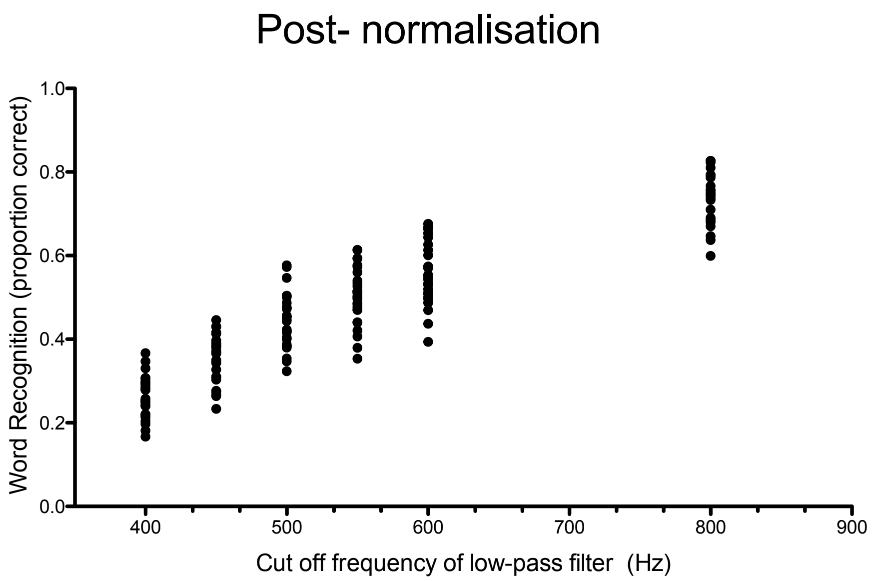


Figure 17. Average word recognition performance of UCAST-FW for all participants for pre-normalisation and post- normalisation conditions, across the six tested levels of low-pass filtering. A circle represents each participant.

The average intelligibility scores across each subject for each level of filtering was assessed in both pre- and post-normalisation conditions, displayed in Figure 17 (each person is represented by a circle). The primary difference between the graphs is not differences in mean, rather the spread of scores around the mean. A tighter cluster of participant performance is seen in the post-normalisation condition, indicating less variability inter-subject variability following normalization. Additionally, a calculation of the coefficient of variation was calculated to quantify the amount of deviation of scores from the mean. Pre-normalisation CV indicates a greater percent deviation from the mean than does pre-normalisation conditions at all levels of filtering, with the exception of 600 Hz. Pre-normalisation conditions show 25.74%, 22.31%, 16.37%, 14.83%, 12.8%, and 12.41% variation from individual participant scores from the mean, for filter frequencies of 400, 450, 500, 550, 600 and 800 Hz, respectively. Post-normalisation conditions show 18.51%, 15.64%, 14.32%, 13.14%, 12.98%, and 8.41% variation from individual participant scores from the mean, for filter frequencies of 400, 450, 500, 550, 600 and 800 Hz, respectively. All CV values are given is Table 6. This indicates a tighter range, with less variation of participant scores for the post-normalisation conditions relative to the pre-normalisation conditions.

Table 6. The coefficient of variation for the word recognition performance among all tested participants for the open set test.

Filter Frequency (Hz)	**Pre-normalisation (%CV)** n= **30**	**Post-normalisation (CV%)** n = **31**
400	26.74	18.51
450	22.31	15.64
500	16.37	14.32
550	14.83	13.14
600	12.8	12.98
800	12.41	8.41

3.1.6 Outliers

Outliers identified in the pre-normalisation test condition were compared to their post-normalisation counterparts to assess if there was any improvement in SRT and slope values following normalisation. Pre-normalisation outliers for SRT are tree, food, head, and meat, and outliers for slope are head and meat. Following normalisation, the only word that shows improvement is food, as displayed in Figure 18. The remaining words present as outliers in both pre- and post-normalisation conditions and show no improvement following normalisation, as displayed in Figure 19. These words are immune to the normalisation process and should thus be considered for removal from the UCAST-FW word list due to their poor performance.

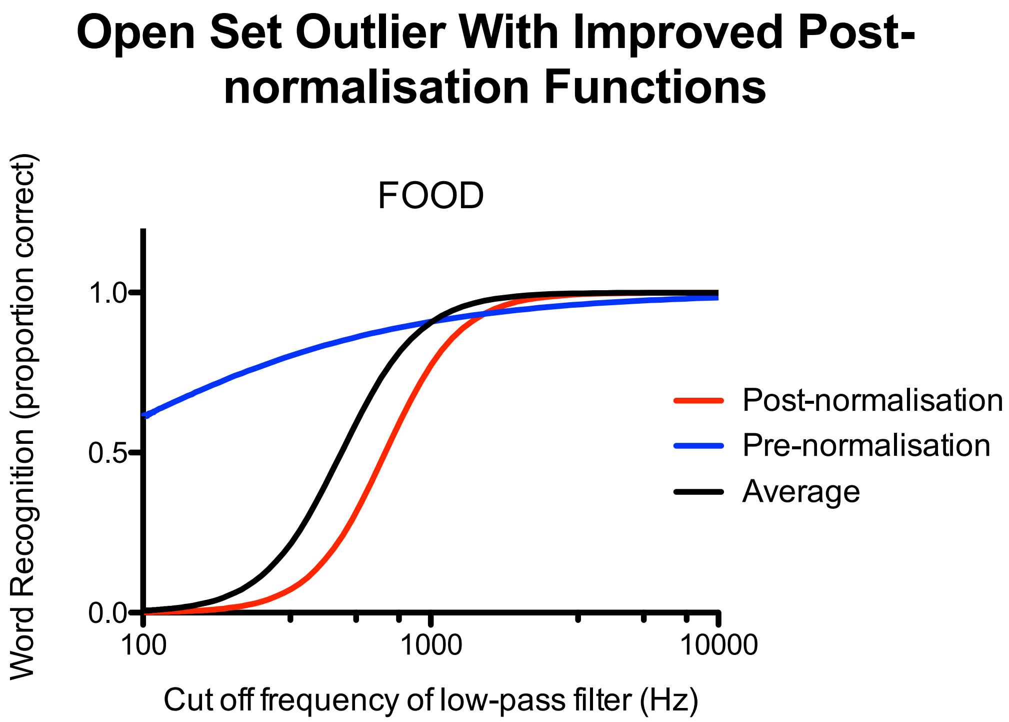


Figure 18. Psychometric functions for the pre-normalisation outlier food. The blue line represents pre-normalisation performance, the red line represents post-normalisation performance, and the black line represents the average. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter.

Words to be Considered for Exclusion from the Open Set UCAST-FW Word List

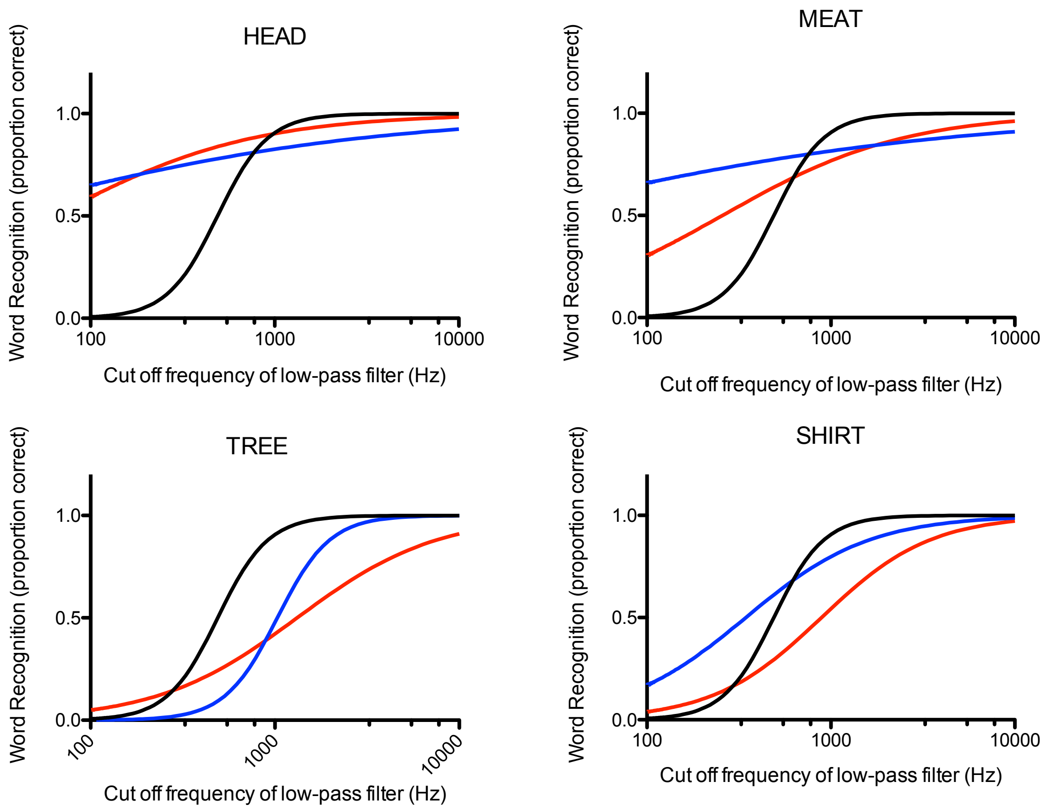


Figure 19. Psychometric functions for words head, meat, tree, and shoe. The blue line represents pre-normalisation performance, the red line represents post-normalisation performance, and he black line represents the average. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter.

3.2 Closed set

3.2.1 Word recognition performance

The effects of the cut-off frequency of the low-pass filter were assessed as a function of overall word recognition scores for both pre- and post- normalisation test conditions. For closed set testing, word recognition performance was scored as either a correct or incorrect response, represented by a one or zero, respectively. Scores were averaged over the number of participants for each condition. The percent (%) of correct word recognition for a particular cut-off filter frequency (200, 300, 350, 400 and 500 Hz) was averaged across all 50 words in the UCAST-FW for all tested participants.

The mean word recognition scores for pre- normalisation testing were 41.5% (*SD* = 14.2), 49.5% (*SD* = 14), 55.1% (*SD* = 16.7), 57.7% (*SD* = 17.5), and 82.9% (*SD* = 20.3) for cut-off filter frequencies of 200, 300, 350, 400, and 500 Hz. The mean word recognition scores for post-normalisation scores were 42.5% (*SD* = 13.3), 43.8% (*SD* = 13.5), 47.3% (*SD* = 12.5), 52% (*SD* = 14.1), and 77.8% (*SD* = 18.7) for the cut-off filter frequencies of 200, 300, 350, 400, and 500 Hz. Results are displayed in Figure 20.

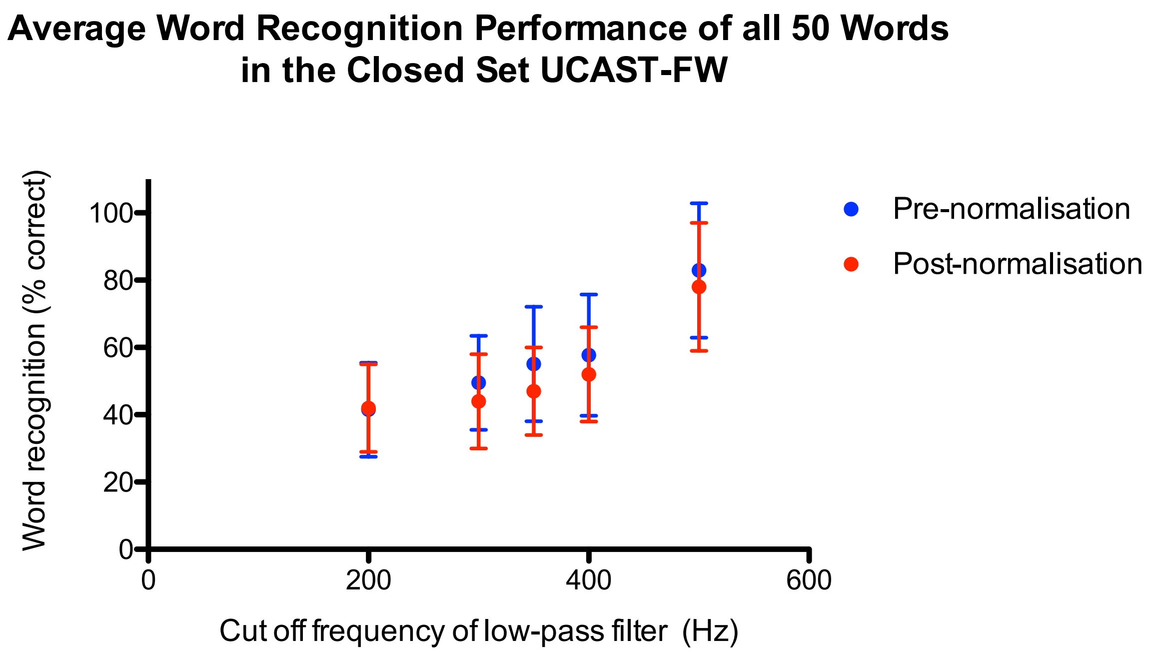
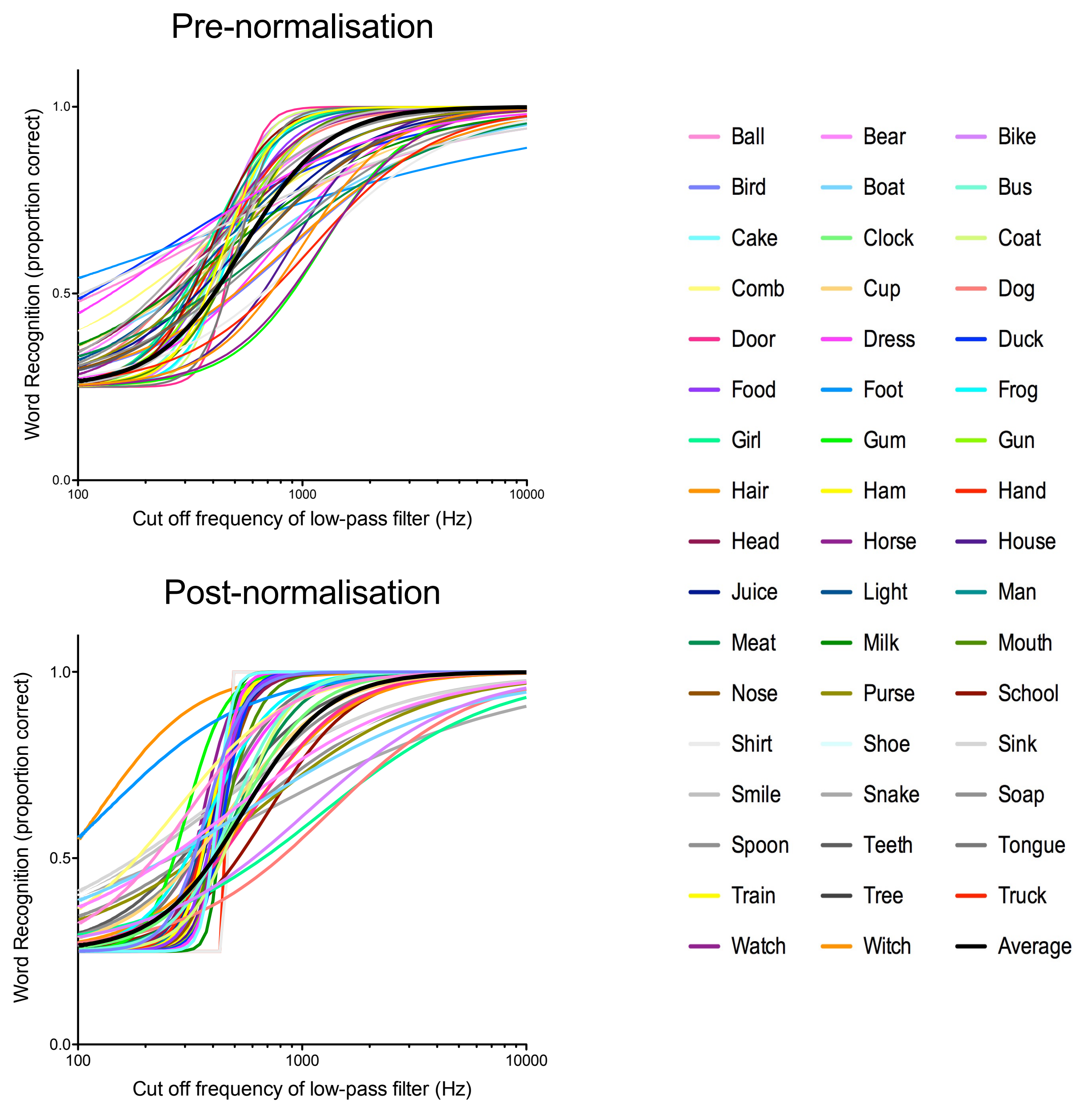


Figure 20. Average word recognition performance across all tested participants and filter frequencies (200, 300, 350, 400, and 500 Hz) for pre-normalisation (blue, *n* =30) and post-normalisation (red, *n* = 31). Word recognition performance for each cut-off filter frequency is averaged across all 50 words in the UCAST-FT word list. Error bars represent standard deviation.

3.2.2 Psychometrically Derived Word Recognition Performance

Following the collection of raw data, logistic regression was used to obtain regression slope and SRT for each word in the UCAST-FW closed set paradigm, for both pre- and post-normalisation testing conditions. Figure 21 displays the psychometric curves for all words in the UCAST-FW test. Intercept values for each word were calculated to derive the cut-off frequency of the LPF that generated SRT values at 62.5% (half-way point between chance performance of 25 % and 100 %) word recognition performance. SRT and slope values for each word are displayed in Table 7. Overall, the psychometric functions for the pre-normalisation test generated thresholds that ranged from 249.1 Hz to 1224.4 Hz. (*M* = 466.3 Hz, *SD* = 247.9). Slope functions range from 33.3 to 60.53 (*M* = 47.1 *SD* = 5.9). For the post-normalisation conditions, the psychometric function generated thresholds that ranged from 123.7 Hz to 1335.8 Hz (*M* = 436 Hz, *SD* = 216.8). Slope functions range from 37.4 to 90.42 (*M* = 53.4, *SD* = 11.1).

Normalisation of the Closed Set UCAST-FW Word List



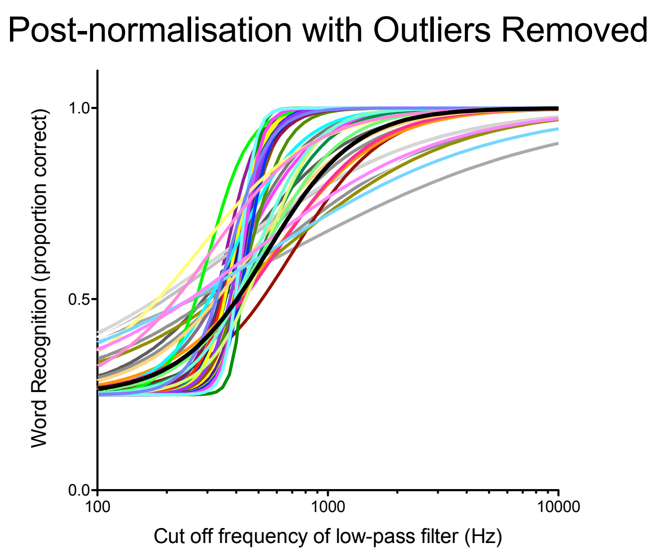


Figure 21. Pre- and post- normalisation psychometric curves and post-normalisation curves with significant outliers removed, for the all words in the UCAST-FW closed set test. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter.

Table 7. Psychometrically generated SRT and slope values for each word in closed set UCAST-FW for pre- and post- normalisation testing conditions. Outliers (discussed below) are shown in grey

	Pre- normalisation		Post-normalisation	
Word	**SRT**	**Slope**	**SRT**	**Slope**
Ball	316.8	37.2	311.6	47.2
Bear	366.1	45.4	478.8	41.2
Bike	421.5	55.6	1059.0	42.7
Bird	892.9	42.8	375.5	56.9
Boat	609.5	40.1	553.1	39.3
Bus	474.4	55.4	417.4	66.5
Cake	437.9	53.6	488.8	52.8
Clock	411.6	52.7	524.6	49.9
Coat	460.5	57.2	528.9	53.7
Comb	357.2	41.5	272.7	46.0
Cup	507.8	41.4	471.6	47.1
Dog	370.4	47.5	1335.8	43.5
Door	490.7	60.5	599.1	47.7
Dress	755.2	45.8	426.0	64.1
Duck	445.4	51.2	416.1	59.8
Food	249.1	39.1	447.6	61.8
Foot	258.1	33.2	132.9	43.7
Frog	483.0	54.7	388.5	51.4
Girl	375.9	51.5	1251.6	41.5
Gum	1224.5	48.5	307.2	56.8
Gun	459.1	50.2	445.5	61.6
Hair	518.8	51.5	393.8	60.1
Ham	872.3	42.7	587.9	47.0
Hand	436.0	50.0	394.8	58.3
Head	390.0	44.1	396.8	56.8
Horse	392.3	48.6	422.6	63.3
House	893.0	48.9	435.7	61.3
Juice	534.1	44.7	406.2	57.8
Light	432.5	44.8	384.1	57.8
Man	393.5	51.2	401.4	59.1
Meat	729.9	41.1	512.1	51.8
Milk	479.4	41.6	440.1	68.2
Mouth	517.7	52	456.9	57.3
Nose	472.1	45.4	620.5	41.8
Purse	604.7	44.5	436.3	59.4
School	391.3	52.8	707.3	48.5
Shirt	416.4	40.4	443.2	39.9
Shoe	1111.1	43.5	463.9	90.4
Sink	282.8	36.9	359.2	40.7
Smile	483.3	50.7	381.7	41.0
Snake	344.4	45.2	675.1	37.4
Soap	404.1	45.8	565.2	41.7
Spoon	731.6	42.8	486.1	46.2
Teeth	518.7	57.5	412.8	51.2
Tongue	601.0	44.8	419.4	46.8
Train	450.0	54.1	417.3	61.3
Tree	1111.2	44.9	460.7	89.9
Truck	1212.4	47.9	362.5	59.2
Watch	992.7	48.1	123.7	47.4
Witch	278.5	40.7	423.2	50.6
Average	547.3	47.05	484.9	53
Average (no outliers)	29.7	47.1	445.7	54.6
Min	249.1	33.3	123.7	37.4
Max	1224.5	60.5	1335.8	90.4
Range	975.3	27.3	1212.1	53.1
SD	247.1	5.9	216.8	11.1
SD (no outliers)	221.67	6.22	82.3	11.32

3.2.3 Departures from Normality- SRT

The distributions of SRT values were tested for normality and homogeneity of variance for both pre- and post- normalisation testing conditions. Pre-normalisation SRT values show a positive skew to the right, with a value of 4.24, and a normal kurtosis value of 1.909. Additionally, a Shapiro-Wilks test was used to test the assumption of normality ($W = 0.824$, $p < .05$). The distributions of pre –normalisation SRTs do not meet the assumption of normality and show a skewed distribution to the left.

Post-normalisation SRT values with outliers included show a positive skew to the right, with a value of 7.136. SRT values also display a large positive kurtosis value of 11.298. Additionally, a Shapiro-Wilks test was used to test the assumption of normality ($W = 0.72$, $p < .05$). The distribution of post-normalisation SRTs with outliers included, do not meet the assumption of normality, and show a leptokurtic distribution skewed to the right, indicating that the data set is clustered and has a heavier tail distributed to the right than a normal distribution. A Shapiro-Wilks test was used to examine the assumption of normality for the distribution of SRT values in post-normalisation conditions, with the removal of significant outliers ($W = 0.9561$, $p = .063$). The removal of significant outliers following normalisation generated a distribution of SRTs that meet the assumptions of normality, with a skewness of 0.196 and kurtosis of 1.224.

The distribution of SRT values, including outliers, for both pre- and post- normalisation conditions are displayed in a box plot in Figure 22. The highest SRT values indicate those words that require a higher LPF for 62.5% recognition (half way between 25% chance performance and 100%), and therefore more speech information is required for correct word recognition, making words with the highest SRT values commonly more difficult. The opposite is true for words with the lowest SRT values. Pre-normalisation words with the highest SRT values, indicating words that are more commonly difficult are gum, truck, tree, shoe, and watch. The lowest SRT values, indicating words that are considered easier are food, foot, witch, sink, and ball. An outlier is defines as a value that is either: 1) <25^{th} percentile – 1.5 x interquartile range or 2) > 75^{th} percentile + 1.5 x interquartile range. Pre-normalisation outliers for the closed set list are: gum (SRT = 1224.5 Hz), truck (SRT = 1212.4 Hz), tree (SRT = 1111.2 Hz), shoe (SRT = 1111.1 Hz), and watch (SRT = 992.7 Hz). Post-normalisation words with the highest SRT values are dog, girl, bike, school, and snake. Lowest SRT values are watch, foot, comb, gum, and ball. From this list of extreme values, six were determined to be outliers: dog (SRT = 1335.8 Hz), girl (SRT = 1215.6 Hz), bike (SRT = 1059 Hz), school (SRT = 707.3 Hz), foot (SRT = 32.9 Hz), and watch (SRT = 123.7 Hz). A list of the extreme values is given in Table 8.

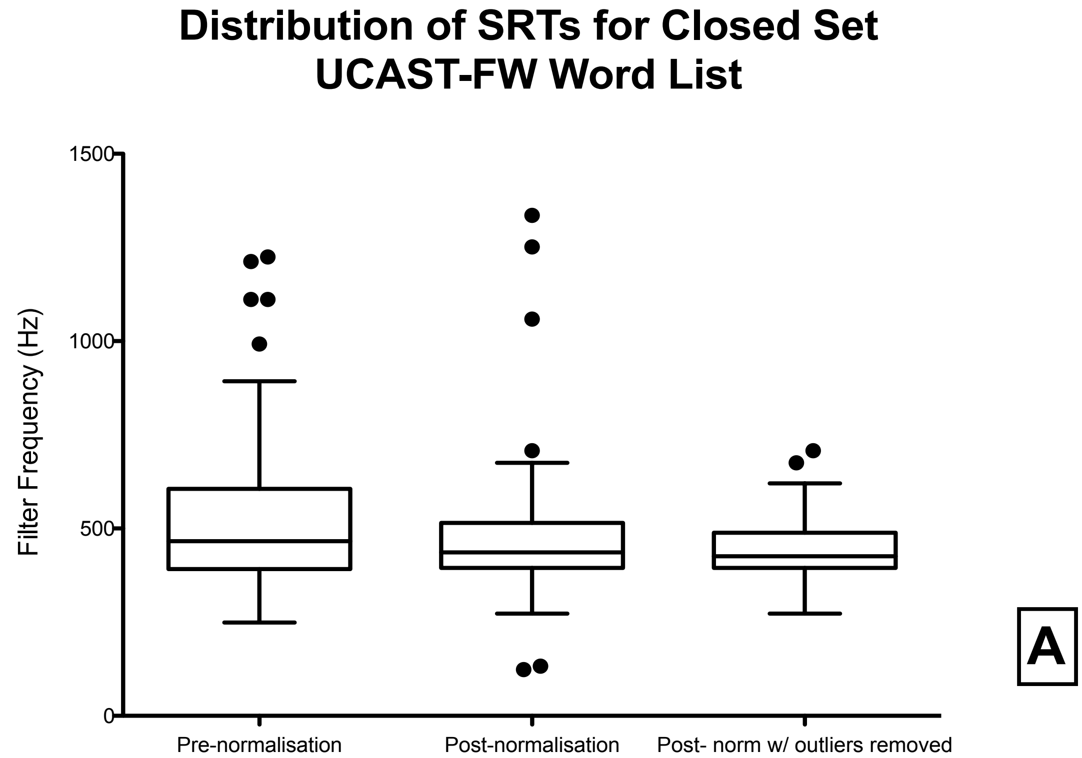


Figure 22. Boxplot of the medians (with interquartile range and outliers) of the distribution of SRT and slope values for open set testing. Distributions of pre- and post normalisation, and post-normalisation with outliers removed are given for both SRT and slope.

Table 8. Five highest and five lowest SRT values are given for pre- and post- normalisation conditions for the closed set test. Outliers (discussed below) are shown in grey.

Extreme SRT Values- Closed Set						
Pre-norm		**Word**	**LPF (Hz)**	Post-norm	**Word**	**LPF (Hz)**
Highest	1	Gum	1224.5	1	Dog	1335.8
	2	Truck	1212.4	2	Girl	1215.6
	3	Tree	1111.2	3	Bike	1059
	4	Shoe	1111.1	4	School	707.3
	5	Watch	992.7	5	Snake	675.1
Lowest	1	Food	249.1	1	Watch	123.7
	2	Foot	258.1	2	Foot	32.9
	3	Witch	278.5	3	Comb	272.7
	4	Sink	282.8	4	Gum	307.5
	5	Ball	316.8	5	Ball	311.6

3.2.4 Departures from Normality- Slope (%/octave)

The distributions of slope values were tested for normality and homogeneity of variance for both pre- and post- normalisation testing conditions. Pre- normalisation slope values show a normally skewed distribution with a value of 0.21, and a kurtosis value of 0.601 Additionally, a Shapiro-Wilks test was used to test the assumption of normality (W = 0. 990, p < .05). The distribution of pre –normalisation slope values do not meet the assumption of normality, and show a platykurtic distribution of values.

Post-normalisation slope values show a positive skew to the right, with a value of 3.934, and a kurtosis value of 4.478. Additionally, a Shapiro-Wilks test was used to test the assumption of normality (W = 0.888, p = .000). The distribution of post–normalisation slope values meet the assumption of normality, and a show a tighter cluster of values than the pre-normalisation condition.

The distribution of slope values for pre- and post-normalisation conditions are displayed in a box plot in Figure 22. The highest slope values indicate words that have a steeper slope. Conversely, words with lower slope values indicate words that have a shallower slope. Pre-normalisation test words with the steepest slopes are door, teeth, coat, bike, and bus. Additionally, the words with the shallowest slope are foot, sink, ball, food, and boat. From this list of extreme slope values, none were determined to be outliers.

Post-normalisation words with the steepest slopes are shoe, tree, milk, bus, and dress. Conversely, words with the shallowest slope are: snake, boat, shirt, sink, and smile. From this list of extreme values, two were identified to be outliers, which are shoe (slope = 90.4 %/octave), tree (slope = 89.9 %/octave). A list of extreme values is displayed in Table 9.

Table 9. Five highest and five lowest slope values are given for pre- and post- normalisation conditions for the closed set test. Outliers (discussed below) are shown in grey.

Extreme slope Values- Closed Set						
Pre-norm		**Word**	**%/Octave**	Post-norm	**Word**	**%/Octave**
Highest	1	Door	60.5	1	Shoe	90.4
	2	Teeth	57.5	2	Tree	89.9
	3	Coat	57.2	3	Milk	68.2
	4	Bike	55.6	4	Bus	66.45
	5	Bus	55.4	5	Dress	64.1
Lowest	1	Foot	33.2	1	Snake	37.4
	2	Sink	37	2	Boat	39.3
	3	Ball	37.2	3	Shirt	40
	4	Food	39.1	4	Sink	40.7
	5	Boat	40.1	5	Smile	41

3.2.5 Comparative analysis

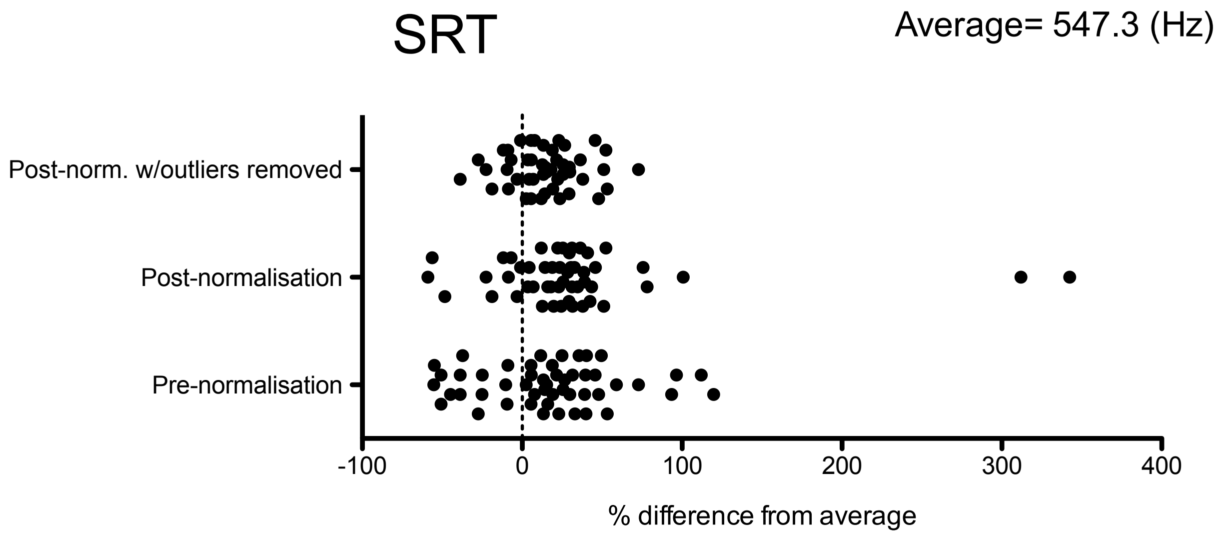
A Related Samples Sign Test indicated a significant difference between pre- and post-normalisation conditions for the closed set SRT (Z = -2.101, p = .036) and slope (Z = -1.838, p = .046). Figure 22 compares the median (with interquartile ranges) and range of SRT and slope values for both pre- and post- normalisation compares the median (with interquartile ranges) and range of SRT and slope values for pre- and post- normalisation conditions. On observation, there is a significant reduction in the spread of SRT values in the post-normalisation condition relative to the pre-normalisation condition. Additionally, a smaller interquartile range between pre-normalisation (IQR = 213.8), and post-normalisation (IQR = 120.7) is observed, indicating a reduction in the variation of SRT values. The opposite is true for slope values, with an increase in the IQR from pre-normalisation (IQR = 8.7) to post-normalisation (IQR = 13.4). This indicates that the normalisation process has increased the variability among slope values for the closed set.

In order to estimate the deviation of the psychometric functions for each word from the average function, the difference in SRT and slope values from the average for a given word was calculated and given as a percent difference, displayed in Figure 23. In general, post-normalisation SRT and slope values show a smaller deviation from the average than pre-normalisation values.

The percent of deviation from the average for each word was calculated and is visualized in Figure 23. Qualitative analysis shows an improvement in the SRT values in twenty-one of the fifty words closer to the average performance, relative to pre-normalisation testing. Additionally, nineteen words (bike, bird, bus, cup, dog, duck, frog, hair, hand, juice, light, milk, mouth, nose, purse, teeth, tongue, train and watch) show a further deviation from the average in the post-normalisation condition. No change was observed in six words (ball, dress, gun, head, man and truck) between pre –and post- normalisation conditions. Of these, six words the words dress and truck shows a change from a negative deviation, to a positive; however, the overall percent deviation remains the same. For example, the word truck showed -54% deviation from the average in pre-normalisation conditions and a 50% deviation in the post-normalisation.

Qualitative analysis shows an improvement in the slope values in nineteen of the fifty words closer to the average performance, relative to pre-normalisation testing. Additionally, twenty-six words (bear, bike, bus, dog, dress, duck, girl, gum, gun, hair, hand, head, horse, house, juice, light, man, mouth, nose, purse, smile, snake, soap, train, tree and truck) show a further deviation from the average in the post-normalisation condition. No change was observed in two words (cake and watch) between pre –and post- normalisation conditions, and three words displayed a large deviation from the average in both the pre- and post- normalisation conditions (boat, shoe, and sink). A table of values for the percent of deviation for each word is in Appendix D.

Percent Difference of all Words in the Closed Set UCAST-FW from the Average



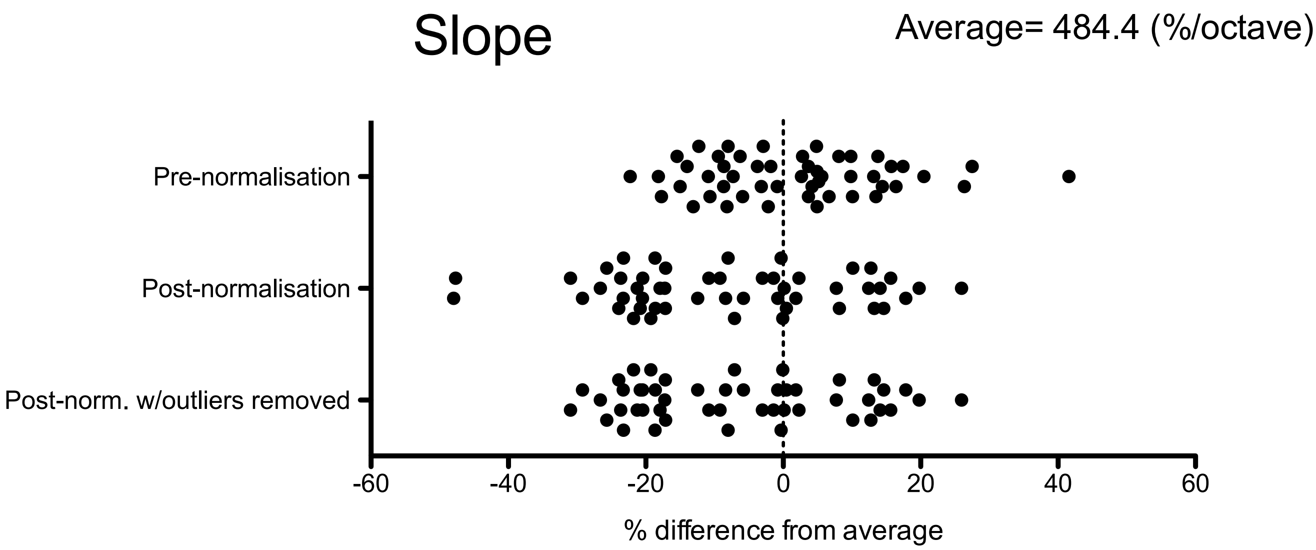
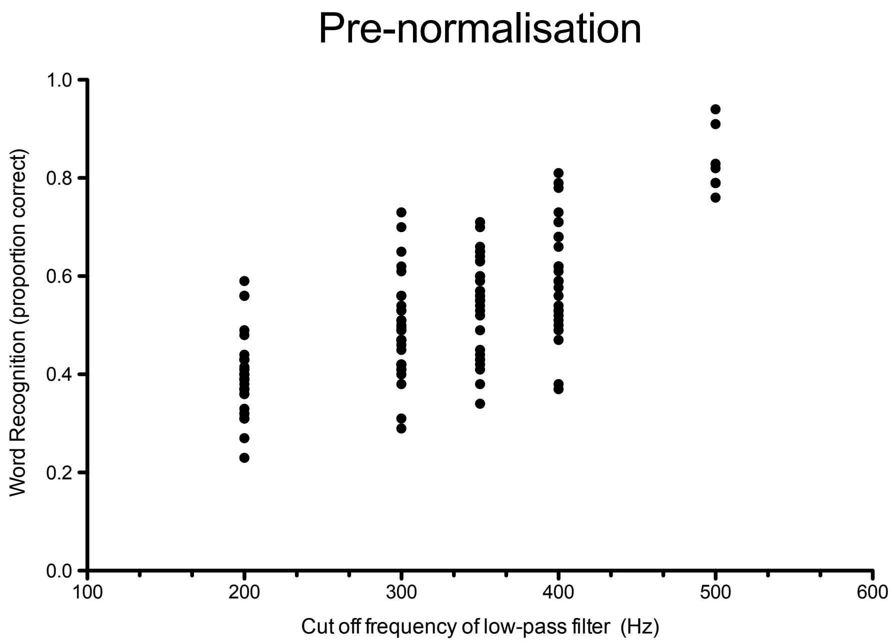


Figure 23. The percent deviation from the average performance for the SRT and slope of each word in the UCAST-FW. The dashed line represents no deviation from the average. Negative values indicate values greater than the average. A circle represents each word in the UCAST-FW test.

Effects of Normalisation on Average Intelligibility Scores Across Each Participant for Closed Set Testing



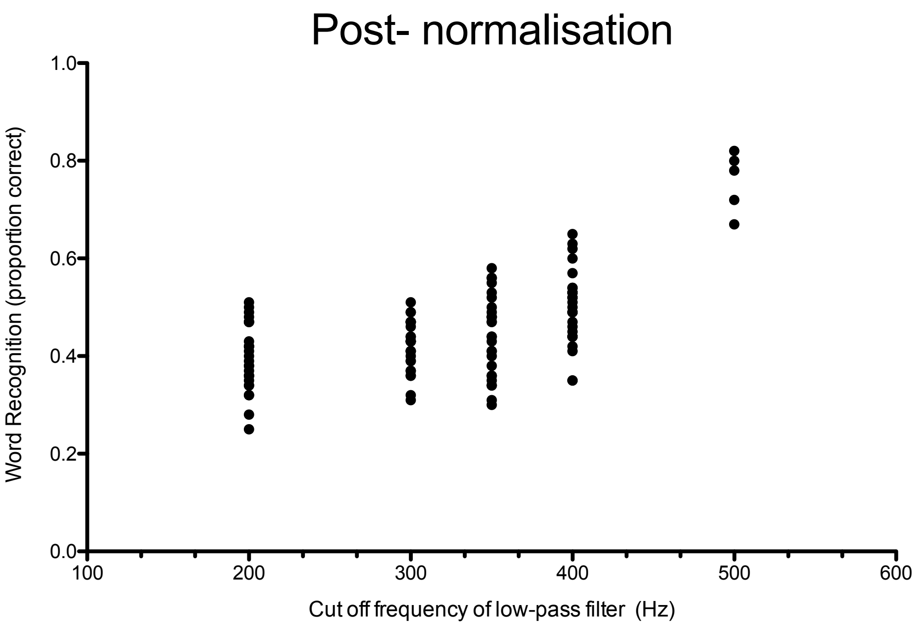


Figure 24. Average word recognition performance of UCAST-FW for all participants for pre-normalisation and post- normalisation conditions, across the 5 tested levels of low-pass filtering. A circle represents each participant.

The average intelligibility scores across each subject for each level of filtering was assessed in both pre- and post-normalisation conditions, displayed in Figure 24 (each person is represented by a circle). In both graphs, the mean increases as filter frequency increases. The primary difference between the graphs is not differences in mean, rather the spread of scores around the mean. A calculation of the coefficient of variation was calculated to quantify the deviation of participant scores from the mean. Pre-normalisation CV indicates a greater percent deviation from the mean than does pre-normalisation conditions at all levels of filtering, with the exception of 600 Hz. Pre-normalisation conditions show 25.74%, 22.31%, 16.37%, 14.83%, 12.8%, and 12.41% variation from individual participant scores from the mean, for filter frequencies of 400, 450, 500, 550, 600 and 800 Hz, respectively. Post-normalisation conditions show 18.51%, 15.64%, 14.32%, 13.14%, 12.98%, and 8.41% variation from individual participant scores from the mean, for filter frequencies of 400, 450, 500, 550, 600 and 800 Hz, respectively. This indicates a tighter range of participant scores for the post-normalisatiion conditions relative to the pre-normalisation conditions.

Table 10. The coefficient of variation for the word recognition performance among all tested participants for the closed set test.

Filter Frequency (Hz)	**Pre-normalisation (CV %)** n= 30	**Post-normalisation (CV%)** n= 31
200	20.17	17.09
300	20.49	12.36
350	17.74	18.4
400	18.3	15.26
500	7.38	8.15

3.2.6 Outliers

Outliers identified in the pre-normalisation test condition were compared to their post-normalisation counterparts to assess if there was any improvement in SRT and slope values following normalisation. Pre-normalisation outliers for SRT are truck, gum, shoe, tree, and watch. There were no pre-normalisation outliers for slope values. Following normalisation, there is an observable improvement in the psychometric functions for the words tree, shoe, and door. Additionally, there is a reduction in the percent deviation from the average performance in these words in the post-normalisation condition.

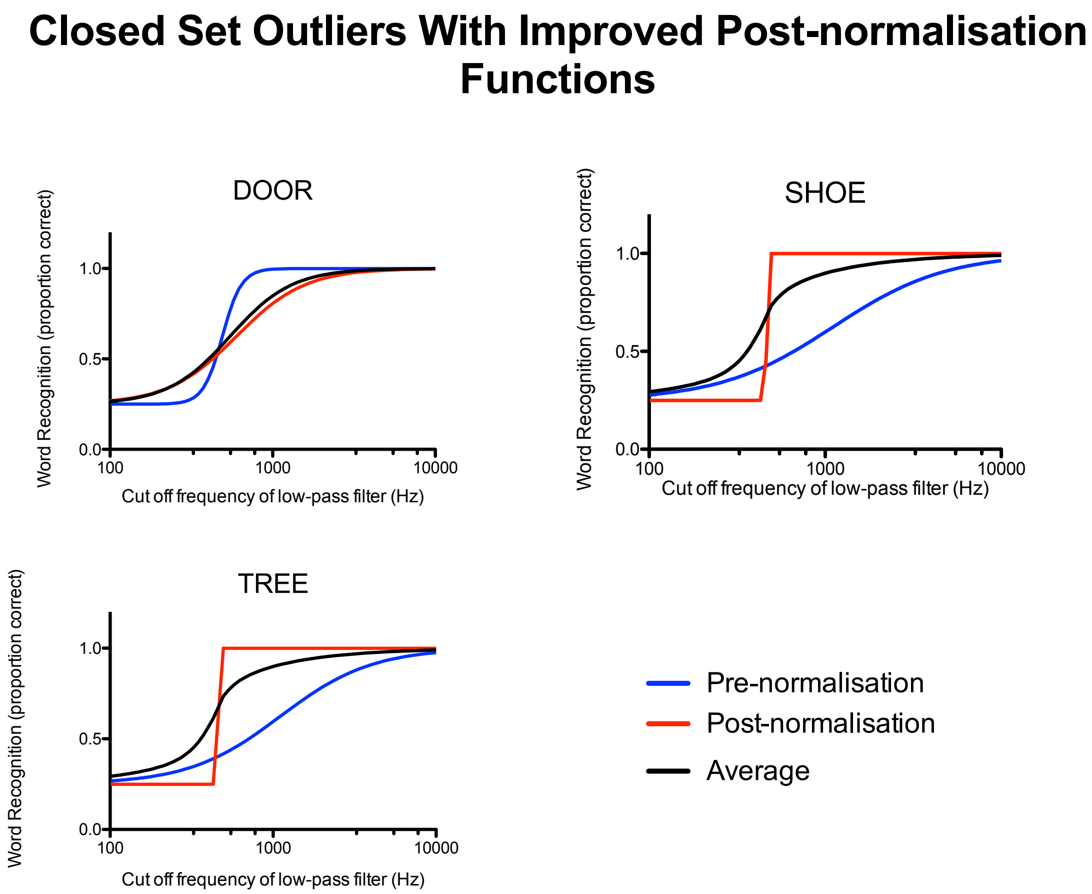


Figure 25. Psychometric functions for the pre-normalisation outliers door, shoe, and tree. The blue line represents pre-normalisation performance, the red line represents post-normalisation performance, and the black line represents the average. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter.

Outliers that are present in both testing conditions, and show large deviations from the average are the words tree, shoe, and watch. These words have poor performance in the pre-normalisation conditions, and are immune to the normalisation process. Additionally, new outliers appear following normalisation, these words are: dog, girl, bike, school, and foot, and do not appear as outliers in the pre-normalisation conditions.

Words to be Considered for Exclusion from the Closed Set UCAST-FW Word List

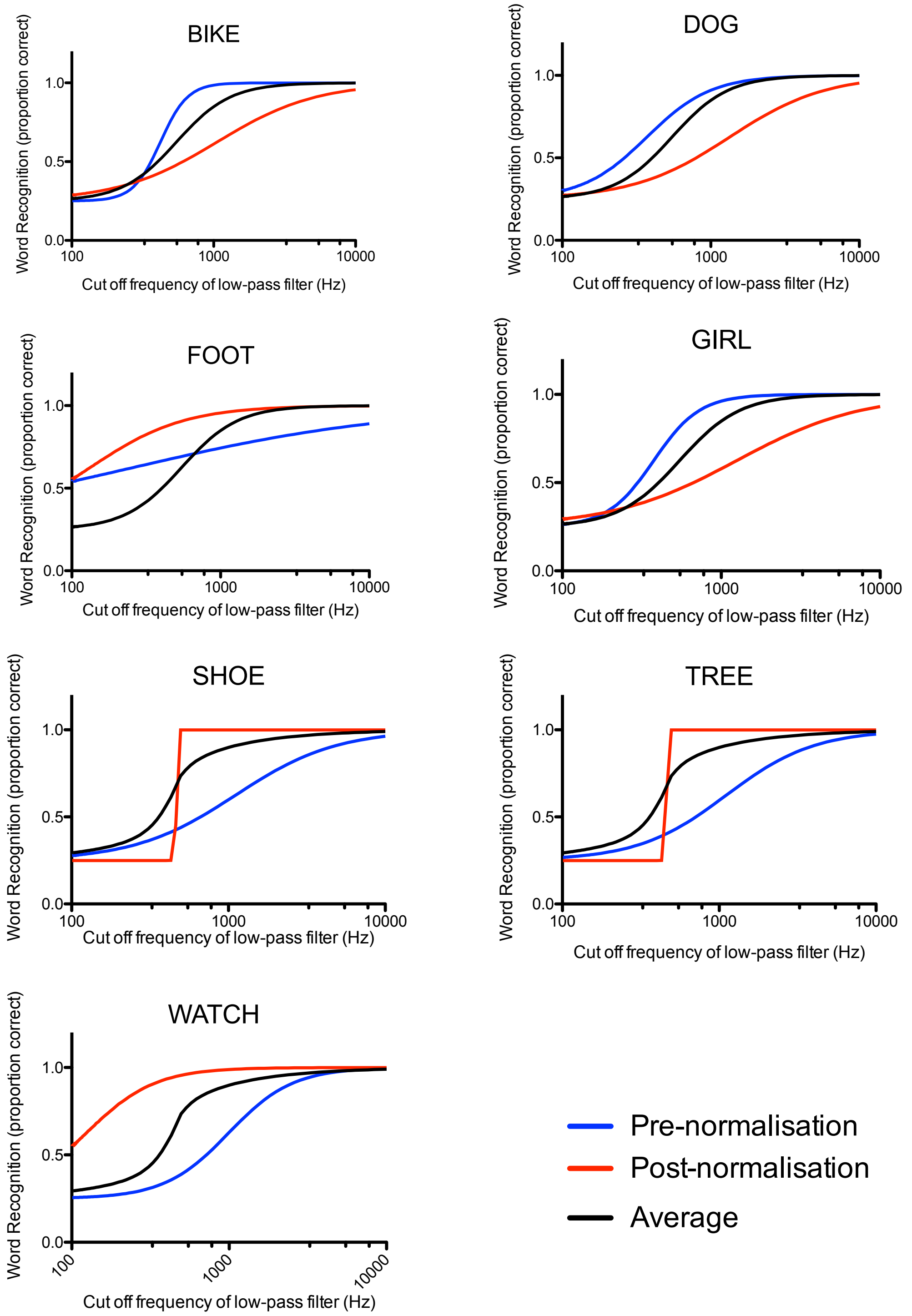


Figure 26. Psychometric functions for words bike, dog, foot, girl, shoe, tree, and watch. The blue line represents pre-normalisation performance, the red line represents post-normalisation performance, and the black line represents the average. Word recognition performance is displayed as a function of the cut-off frequency of the low-pass filter.

Table 11. Normalised levels of filtering for all words as a function of the pseudo-levels of filtering (400, 500, 550, 600 and 800 Hz) for open set testing.

Word	400	450	500	550	600	800
Ball	312.9	363.2	415	468.2	522.7	752.3
Bear	386.2	449.1	514.1	580.9	649.5	939.2
Bike	417.5	490.7	566.9	646	727.9	1079.7
Bird	168.5	288.8	467.7	723.4	1077.2	4017.6
Boat	410.1	485	563.6	645.7	730.9	1101.5
Bus	369.1	419.9	471.2	522.9	575.1	787.8
Cake	306.7	363.8	424	486.9	552.4	838.7
Clock	382.2	409	434.6	459.1	482.7	569.7
Coat	420.1	488.7	559.4	632.2	707	1022.8
Comb	477.5	555.3	635.5	718	802.6	1160.1
Cup	329.7	396.2	467	541.9	620.7	972.3
Dog	351.8	386.6	420.6	454	486.7	612.8
Door	531.6	590.3	648.3	705.7	762.4	984.7
Dress	414.9	530.5	660.9	806.3	966.8	1762
Duck	429.7	446.6	462.3	476.9	490.6	539
Food	150	166.1	193.2	221.4	250.7	378.2
Foot	526	580	633.1	685.3	736.7	935.7
Frog	368.5	399.5	429.5	458.5	486.7	592.8
Girl	372.8	417	461	504.8	548.4	721.2
Gum	356.8	388.5	419.2	449.1	478.2	588.7
Gun	346.2	372.2	397	420.9	443.9	529.6
Hair	381.3	425.3	469	512.4	555.5	725.4
Ham	314.2	343.6	372.2	400.2	427.5	531.8
Hand	266.6	286	304.6	322.5	339.7	403.5
Head	352.1	414	478.5	545.6	615	913.5
Horse	343.8	374.8	405	434.3	463	571.8
House	405.2	479.7	557.9	639.6	724.6	1094.4
Juice	569.3	657.7	748.2	840.9	935.5	1330.6
Light	350.4	384.6	418	450.7	482.7	605.9
Man	343.8	358.4	371.9	384.6	396.6	438.9
Meat	352.1	414	478.5	545.6	615	913.5
Milk	247.8	323.2	409.9	508.3	618.5	1183.6
Mouth	384.5	452	522.4	595.5	671.1	996.4
Nose	687.4	738.9	788.1	835.6	881.3	1051.3
Purse	428.3	522.2	623.6	732.2	847.7	1376
School	256.5	317.4	384	456.2	533.9	898.1
Shirt	233.5	290.5	353.2	421.5	495.3	844.6
Shoe	670.9	768.7	868.2	969.2	1071.7	1494.3
Sink	399.3	531.7	686.8	865.8	1069.7	2152
Smile	339.5	368.6	396.7	424.1	450.7	551
Snake	321	348	374.1	399.3	423.9	516.3
Soap	365.9	425.4	486.9	550.1	615	888.9
Spoon	270.4	373.4	498.3	647	821.1	1805.5
Teeth	352.1	414	478.5	545.6	615	913.5
Tongue	516.4	564.3	610.9	656.3	700.7	870.1
Train	532.9	596.2	659.1	721.8	784.1	1031.3
Tree	352.1	414	478.5	545.6	615	913.5
Truck	529.3	603.8	679.2	755.5	832.7	1148.3
Watch	351.3	452.1	566.5	694.7	837	1549.7
Witch	193.3	259	336.5	426.4	529.2	1081.2

Table 12. Normalised levels of filtering for all words as a function of the pseudo-levels of filtering (200, 300, 350, 400, and 500 Hz) for closed set testing.

Word	200	300	350	400	500
Ball	150	150	150	172.5	299.8
Bear	150	181.8	219.3	257.9	338.3
Bike	241.7	302.3	329.1	354.2	400.6
Bird	150	209.6	345.3	532.3	1096.6
Boat	150	150	181	286.3	615.9
Bus	291.4	354.2	381.6	406.9	453.1
Cake	218.7	289.7	322.4	353.7	412.8
Clock	195.7	265.3	297.8	329.1	389.1
Coat	295.8	352.1	376.3	398.5	438.6
Comb	150	150	180	226.3	331.5
Cup	150	150	207.3	278	453.8
Dog	150	188.8	226.9	266	347.1
Door	355	401.6	420.9	438.4	469.2
Dress	212.9	332.9	394.6	457.2	584.8
Duck	195.2	273.2	310.4	346.7	417.2
Food	150	150	150	170.8	256.8
Foot	150	150	150	150	238.7
Frog	244	326.6	364.9	401.6	471.5
Girl	172.4	237.1	267.7	297.4	354.4
Gum	262.9	565	755.6	972.1	1480.9
Gun	169.3	257.9	302.6	347.5	171
Hair	205.8	306.5	356.6	406.5	506.1
Ham	150	205.8	334.4	509.2	1028.3
Hand	181	255.9	291.9	327.2	395.8
Head	150	170.2	212.5	257.6	355.2
Horse	150	203.2	244.1	286.1	373.2
House	370.2	461.1	501.3	538.9	608.1
Juice	150	225.2	278.7	335.3	456.5
Light	150	190	236.9	286.7	394.6
Man	171	242.2	276.4	310	375.4
Meat	150	150	234.2	371.4	802.5
Milk	150	163.9	215.7	273.7	407.2
Mouth	228.7	321.7	366.3	409.9	494.6
Nose	150	214.2	264.2	316.9	429.5
Purse	150	234.8	301.2	373.7	535.8
School	204.1	263.1	289.7	315	362.1
Shirt	150	150	179.9	235.5	369.2
Shoe	150	281.4	502.4	830	1920.5
Sink	150	150	150	151	275.4
Smile	196.2	289	334.9	380.4	470.8
Snake	150	165.8	202.3	240.4	320.6
Soap	150	185.1	230.2	278.1	381.3
Spoon	150	211.6	301.1	408.8	681.4
Teeth	364.4	412.6	432.5	450.6	482.4
Tongue	150	254	317.7	385.6	533
Train	222.1	299.7	335.8	370.6	437
Tree	150	343.4	568	878.2	1819.4
Truck	187	543.7	815.8	1159.3	2085.8
Watch	440	508.8	537.7	564	611
Witch	150	150	161.3	196.5	273.5

Chapter Four:

4 Discussion

4.1 Introduction

The purpose of this research was to evaluate the homogeneity of the UCAST-FW word list under conditions of low-pass filtering, in order to create a more valid test for the diagnosis of APD. Speech items that produce similar levels of word recognition performance under conditions of filtering reduce the inter-item variability and improve the inter-patient variability, creating a diagnostic test that is more sensitive to changes in an individual performance that are due to auditory dysfunction. The current UCAST-FW word list has some speech items that become considerably more unintelligible when a low-pass filter is applied, relative to other words in the list at the same levels of filtering. This difference in performance is largely due to the spectral variation of each individual test word. When a low-pass filter is applied, the intelligibility of each speech item depends on the frequency range of key acoustic and phonetic cues, and whether or not these cues are available to the listener at different levels of filtering. The large variance among the spectral content of the individual test items within the UCAST-FW creates a word list that becomes heterogeneous in regards to recognition performance under the same levels of filtering. This creates inherent vulnerabilities within the sensitivity and specificity of a diagnostic test. The UCAST-FW diagnostic test is designed to be administered in an adaptive format, during which the selection of words from the list is random, and the level of filtering is given based on the listener's previous response. For this, we must assume that it is the level of filtering that is producing changes in an individual's performance, not the selection of a given word, and that all words will produce the same level of word recognition performance at a given level of filtering. However, the results of this study show that this assumption is incorrect and different words produce different levels of word recognition performance under the same levels of low-pass filtering. Therefore, as the test currently stands, the selection of words, by chance, could produce results that are better or worse than expected than an individual's true auditory capability.

Within in this study, we created a novel method of equalising the difficulty of the speech items by adjusting the levels of filtering, in attempts to compensate for the spectral variability and the differences in word recognition performance under the same levels of filtering. That is, the normalisation process was developed to ensure that the word list produces similar levels of word recognition performance under the same levels of filtering. It was reasoned that this novel approach to normalisation could achieve greater homogeneity over more common methods, as it attempts to normalise both the SRT and slope of each speech item in the test to become more analogous to the average. Previous methods of normalisation generally only shift the SRT by some specified value, regardless of the psychometric slope of the test word. This new method of normalisation was created to ensure that when the adaptive procedure is applied to the UCAST-FW test, a correct or incorrect response, which results in a corresponding increase or decrease in the level of filtering, is due to the individual's auditory capabilities, and not due to the variability within the word list. The range and distribution of SRT and slope values, as well as subjective verification of the post-normative psychometric functions, relative the average were assessed as the outcome measure of this study.

Within this study, we examined the data to determine whether our novel method of normalisation created a more homogenous word list relative to pre-normalisation conditions. Three aims were developed, each with a set of contributing hypothesis to determine whether these aims were satisfied. The following section reviews the aims and hypotheses with the evidence collated from this research.

4.2 Assement of aim one

Aim 1*: To normalize the difficulty of the UCAST-FW word list by adjusting the level of low-pass filtering, equal to the average pre-normalisation word recognition performance.*

Hypothesis 1: We predicted that there would be a reduction in the spread of distribution of (a) equivalent SRT and (b) slope values following the normalisation process.

Hypothesis 2: We predict that: there will be a shift in the (a) SRT and (b) slope values for all words in the UCAST-FW word list relative to the average pre-normalisation condition following normalisation.

The results of this study support hypothesis one and hypothesis two for both open and closed set paradigms.

4.2.1 Open set

Using the open set paradigm, we examined the ability of the normalisation process to create a less variable distribution of the SRT and slope values across all words in the UCAST-FW word list. We found that following normalistion, the distribution of SRTs was significantly reduced, as indicated by a smaller interquartile range, showing a tighter and less variable distribution of SRT values. Additionally, the pre-normalisation distribution of SRTs did not meet the assumptions of normality, due to the large range and skewness of individual SRTs. However, following normalisation, the variability of spread was reduced, and the distribution met the assumptions for normality. A tighter, more normally distributed range of SRTs following normalisation provides a more homogenous word list in relation to SRTs.

The distribution of slope values shows a less pronounced reduction in spread than the SRT; however, following normalisation we see that the overall percent deviation of individual slope values is closer to the average. As the aim of the normalisation process was to create psychometric functions that are more analogous to the average, the degree of deviation of individual SRT and slope values from the average was assessed. A reduction in the amount of deviation from the average following normalisation indicates that the normalisation process has created psychometric functions that are closer to the average, thereby equalizing the difficulty of the test. Overall, the normalisation process reduced the percent of variation for SRT values from the average from 117.36% to 4.68%, and from 2.09% to -1.07% for slope values. This shows that on average, we have SRT and slope values that are more similar to the average, indicating that the normalisation process is equalising the relative intelligibility of the word list to the average.

The novel aspect of our normalisation process is that it attempts to equalise the difficulty of each word relative to the average performance, rather than some arbitrary subjective value. The average performance was chosen as the point to normalise against, as each word in the test was a contributing factor to the average psychometric function, providing an objective value to normalise against. This means that all words, regardless of difficulty or word recognition performance under conditions of low-pass filtering, contributed to the average psychometric curve, allowing a set point that each individual word could be comparatively assessed against. Additionally, an average value dictates that the level of normalisation should be similar for the most difficult and most easily recognizable words, which may not be the case if we chose another value to normalise against.

4.2.2 Closed set

A similar pattern is seen for the distribution of SRT values in the closed, with a reduction in the interquartile range of SRT values for post-normalisation conditions. Closed set SRTs also meet the assumptions for a normal distribution following normalisation. A normal distribution indicates a more homogenous and less variable data set than the pre-normalisation condition. When assessing the percent of variation from the average for SRT values, we see an increase in the percent difference from the average performance for closed set SRT values from 16.61% in the pre-normalisation condition to 32.98% in the post-normalisation condition, the opposite effect of what we expected to see. On further investigation of this finding, we see a tighter cluster of individual SRT values closer to the zero line (zero indicates no percent change from the average) in Figure 23, indicating that the normalisation process has in fact reduced the percent difference of the majority of words following normalisation, despite an increase in the average percent difference after normalisation. The reason the overall percent variation increased following normalisation is because of the distribution of positive and negative values. A percent difference was calculated for each of the 50 words, giving either a positive value (indicating a SRT smaller than the average), or a negative value (indicating an SRT larger than the average). In the pre-normalisation condition, there is a more even spread of positive and negative values. However, when just reviewing the magnitude of change from the average, with only positive values, the average magnitude of change shows a reduction from 35.7% to 27.1% following normalisation. Overall, we see a tighter congregation of values closer to the average following normalisation, indicating less deviation of the average. In combination, the reduction in spread, interquartile range and percent variation for the SRT values indicates that our novel method of normalisation has increased the homogeneity of SRT values for both open and closed set.

4.3 Assement of aim two

Aim 2: *To determine whether any words in the UCAST-FW were required to be excluded from the word list.*

Hypothesis 1: There will be outliers present in the post-normalisation conditions for SRT and slope function, indicating words that show no improvement following normalisation and will be considered for exclusion from the UCAST-FW word list.

4.3.1 Aim two summary

The second aim of this study was to identify any outliers within the UCAST-FW word list in terms of SRT and slope and decide if these outliers should be excluded from the word list. The results of this study support the exclusion of four words from the open set paradigm, and eight words from the closed set paradigm.

As expected, the normalisation process results in an observable shift in the psychometric functions toward the average for the majority of the words. However, we cannot expect this trend in all words due to the inherent variability in the spectral content of each word in the test. Some words in the test show no change in word recognition performance following normalisation. By increasing or decreasing the level of filtering, not all words will have a corresponding change in word recognition performance. The reason for certain words have little change in auditory perception following normalisation can be categorized into intrinsic and extrinsic factors. One intrinsic factor that affects auditory perception relates to the spectral variation of each individual test word. The addition of a low-pass filter changes the intelligibility of the speech item, dependent on the frequency emphasis of the item. If a word has a particularity high-frequency emphasis, then the amount of important speech information available to the listener will not be available with a low-pass filter, and even with greater cut-off frequencies of the low-pass filter, the presented sound may sound like a nonsense syllable or word. This means that for some words changing the level of filtering will have little effect on word recognition performance unless the cut-off frequency is greater than the frequency range for those key acoustic cues.

One extrinsic factor that affects auditory perception is related to an individual's lexical memory. An observation throughout the open set testing was that participants would repeat the same word when a certain nonsense syllable was heard. For example, of the words beginning with a */m/* syllable (such as meat, milk, man and mouth), the most common response were meat, creating a poorer performance on other words beginning with */m/* and better performance for the word meat. When the speech signal is degraded and only a portion of the speech item is heard, words with similar sounding phonemes are activated in the listener's memory, which they must choose the most likely option. These acoustic-phonetic patterns are processed relative to the similarities of the input signal. This process is explained in the neighborhood activation model (NAM) of auditory word recognition, which describes the process by which a stimulus word is identified in the context of phonetically similar words activated in memory. The idea of the NAM for influencing word recognition supports previous literature by Vitevitch (2003). Here, they showed that participants attempting to identify degraded real words could be influenced to repeat certain words, developing a processing bias through manipulation of the neighborhood density and neighborhood activation model. The present supports the idea that when some words are exposed to a low-pass filter, the listeners may repeat words that are more common in their lexical memory, creating a bias toward lexical processing rather than auditory processing.

Additionally, the exclusion of outliers does not necessarily relate to the efficacy of our novel method of normalisation. For some words that have been recommended for exclusion, we see a large improvement from the pre-normalisation values to the post-normalisation values closer to the average, however, even with the improvement; the values are still outside of the normal range. This indicates that the normalisation process has in fact improved word recognition performance closer to the average, but the intrinsic and extrinsic factors described above are influencing word recognition performance. For example, the word meat shows a pre-normalisation SRT of 15.93 Hz (2972.05% deviated from the average) and a post-normalisation SRT of 257.34 Hz (90.17% deviated from the average). While there is a large improvement in the SRT following normalisation, it still falls outside of the normal range and is thus still considered for exclusion.

4.3.2 Open Set Outliers

Outliers identified in the pre-normalisation conditions are tree, food, head, and meat. Normalisation improved the psychometric function of the word food; however, tree, head, and meat remained outliers for the post-normalisation condition. This indicates that the poor word recognition performance of the words tree, head, and meat in the pre-normalisation condition were not resolved following normalisation, indicating that the normalisation process had no positive influence on word recognition performance of these words. Thus, the UCAST-FW test would benefit from the exclusion of these words in order to create a more homogenous word list. Additionally, some words present as outliers following normalisation, but not during the pre-normalisation condition. This could be due to the reduction in the interquartile range of SRT and slope values following normalisation, resulting in outliers that were not apparent in the pre-normalisation condition due to the increased spread. These words are shirt and teeth. Although the word shirt is an outlier in the post-normalisation conditions, the psychometric curve for the word shirt is not extremely deviated from the average, indicating that the normalisation process still had a significant effect on the words performance relative to the pre-normalisation condition. The word teeth shows similar psychometric functions for both pre-and post-normalisation conditions, which are both largely deviated from the average, again indicating that the normalisation process had no positive influence on word recognition performance.

Based on this, for the open set paradigm, we recommend the exclusion of the words tree, head, meat, and teeth from the UCAST-FW word list.

4.3.3 Closed Set Outliers

In the closed set paradigm, we see a greater proportion of words that fall outside the normal distribution of SRT values, than in the open set paradigm.

Closed set testing yielded a higher number of outliers for SRT values than the open set. Outliers identified in the pre-normalisation condition are door, truck, gum, shoe, tree, and watch. Following normalisation, all words except for shoe and watch showed improved psychometric functions and were no longer considered outliers. Outliers that appear in the post-normalisation condition but not the pre-normalisation condition are dog, girl, bike, school, tree, watch, and foot. These words have psychometric functions that are largely deviated from the average. In order to increase the homogeneity of the UCAST-FW, the individual speech items recommended for exclusion from the closed set paradigm are: bike, dog, foot, girl, shoe, tree, and watch.

4.4 Assement of aim three

Aim 3: *To determine the differences in word recognition performance in the pre- and post-normalisation conditions between the open and closed set paradigms, and determined what paradigm is more suitable for the UCAST-FW word list.*

Hypothesis 1: The closed set results will show a tighter distribution of SRT and slope values, relative to the open set

Hypothesis 2: The closed set paradigm will be more sensitive to changes in word recognition performance with changes in the level of low-pass filtering

Unexpectedly, results of this study indicate that the open set paradigm generates a more homogenous word recognition performance that is more sensitive to the effects of changes in the low-pass filter, than the closed set paradigm.

In general, a closed set format is considered easier than an open set task, due to the high level of chance performance with the closed set, and the wider range of possible responses with the open set format. This difference was compensated for within the experimental testing by having lower levels of cut-off filter frequencies, compared with the open set, and having highly confusable response foils. Additionally, psychometric curves were generated with the origin point being 0.25 for the closed set, indicating that word recognition performance could not fall below chance performance for the closed set analysis. Regardless, the open set format produced more reliable and consistent results relative to the closed set format. Open set testing showed fewer words that had extreme deviations from the average than the closed set paradigm. Additionally, more words showed improved word recognition performance under conditions of filtering following the normalisation process. This indicates that the open set paradigm is more sensitive to the effects of small increases in the low-pass filter frequency on word recognition performance. This was apparent as the open set testing had significantly shallower psychometric curves than the open set. A psychometric curve that is shallow indicates that a small change in filter frequency will elicit a small, and more progressive change in word recognition performance, compared with a steep slope in which a small change in filter frequency elicits are large change in word recognition performance. Thus, a word with a shallower slope is more sensitive to the effects of filter frequency than a steeper slope

Additionally, the open set paradigm is more sensitive to changes in word recognition performance, due to the differences in scoring between the open and closed set paradigm. CVC scoring occurs on a phoneme-by-phoneme basis. This allows for a more precise measure of reception. Therefore, the open set paradigm is likely to give results that are more valid when the adaptive procedure is applied relative to the closed set, as we are more likely to gain a precise measure of an individual's threshold when the test is sensitive to small increases in filter frequency. These differences experienced between the open and closed set paradigms may be the result of the number of response options. In the closed set, we assume chance performance is 25% with four response options. However, chance performance is only relative when the listener is guessing. With closed set, the listener is inclined to give an educated guess, based on the partial availability of acoustic cues. This means that when a particular sound is heard, the can deduce the correct word based on this sound, and eliminate incorrect options that do not contain this sound. This enhances the probability of choosing the correct response, relative to an open set format when no visual prompts are given and the listener is required to rely on lexical memory to deduct the correct word. For these reasons, the open set format is likely to produce more reliable results of an individual's ability to recognize words under conditions of low-pass filtering. The disadvantage of the using the open set format is the age at which children can perform the task. Closed set testing is more children appropriate and can be administered to young children quickly and reliably.

5 Conclusion

Within this study, we created a new method for the normalisation of speech tests and applied it to the UCAST-FW test. This new method offers benefits for creating a more homogeneous word list by taking into consideration not only the SRT but also the psychometric slope of the test word, and can be normalised to some specified standard, which in this case was the average performance. Optimization of the UCAST-FW word list improved the psychometric functions for a large proportion of words in both the open and closed set paradigm. Overall, we see a reduction in variability of psychometric functions within the UCAST-FW word list for both open and closed set paradigms, indicating that the normalisation process created a more homogeneous word list relative to the pre-normalisation conditions. However, some words show no improvement in word recognition performance under conditions of low-pass filtering following normalisation. This difference in performance is largely due to the spectral variation of each individual test word. Additionally, the influence of an individual's lexical memory related to the NAM can change the perception of the test word when the key phonetic cues are partially available. This results in some words that show no improvement in word recognition performance for levels of filtering below the frequency range of key acoustic cues. This concept makes the process of normalisation difficult when filtered words are involved. However, it is reasoned that this novel method of normalisation will provide a more homogeneous word list relative to the more common methods by applying adjustment factors to the level of filtering that assists in compensating for theses spectral variations within each word. Words that show little or no improvement in word recognition following normalisation represent words that are, for either intrinsic or extrinsic reasons, immune to the normalisation process and are therefore recommended for exclusion from the UCAST-FW word list. These words pose a threat the validity and reliability of the UCAST-FW test for the diagnosis of APD. From this study, we recommend the removal of the words tree, food, head, and meat from the open set paradigm, and the words bike, dog, foot, girl, shoe, tree, and watch from the closed set. Further, the results of this study indicate that for a precise and reliable measure of a threshold of an individual's ability to recognize filtered words to assist in the diagnosis of APD, an open set paradigm is preferred over the closed set.

6 Limitations and Future Directions

The present study has a number of limitations involving the participants, the stimuli, and the lack of statistical power. The participants in this study are not representative of the general adult population; the majority of participants were female students from the University of Canterbury department of Communication Disorders, and the remaining students were all from within the University itself. As University students, and in particular audiology students, the language ability of participants and their previous exposure to speech tests, including familiarization of the NUCHIPS is likely to be higher than that of the general population. This makes it difficult to generalise how a random population will perform. To ensure that the present study's selection of participants did not confound results, replication in a more representative adult study is recommended. Additionally, lack of statistical power due to an insufficient number of participants could be resolved by increasing the sample size. For a Related Samples Sign Test (two-tailed) to have enough statistical power, 60 participants would be needed for each testing condition, which is double the number of participants used in this study. However, due to funding, and time restrictions 30 participants was deemed sufficient to gain a general understanding of word recognition performance under different levels of filtering, and about the effects of the normalisation process.

Another aspect of the UCAST-FW test that could be considered for revision is the current use of a recording featuring an Australian speaker. Future research should also look at the influence of a male Australian speaker on auditory repetition. The NUCHIPS word list, from which the UCAST-FW word list was developed, was originally spoken in an American English dialect. Thus, the response foils for the closed set are not likely to translate into confusable response options in an Australian dialect. The diagnostic efficacy of the UCAST-FW may benefit from the development of a new word list and picture responses, which contain response foils appropriate for the New Zealand dialect, such as the one produced by Murray (2012). Additionally, the difference between the Australian speaker of the UCAST-FW word list and the common New Zealand accent was apparent and caused some confusion over word recognition under conditions of filtering. This further reinforces the need for New Zealand appropriate word lists, and a speaker of native New Zealand English for more sensitive diagnostic efficacy. Another flaw within the UCAST-FW stimuli is the method of scoring. For the open set paradigm, scoring is performed as a proportion of correct word based on CVC. However, the majority of words in the UCAST-FW word list are not CVC words. Words such as tree make appropriate scoring difficult which may contribute to outliers and extreme values.

Moreover, within this study, the intelligibility of the test items was measured on adult listeners with normal hearing who are native speakers of NZ English. The next steps would be to investigate the applicability of the normalised UCAST-FW word list on New Zealand children. Additionally, future research needs to assess the benefit of the normalised UCAST-FW test on children and the sensitivity to distinguish between children with and without APD. This should be viewed in both the open and closed set paradigms, as this study identified significant variations between the performance of words under open and closed set testing conditions. Within this study, it was also identified that some words were immune to the normalisation process, and this is likely due to the spectral variability within each word. To further improve the normalisation process, each word could be assessed to determine the important frequency ranges for correct word identification, and the optimal position of the low-pass filter for achieving various proportions of word recognition performance. This could then be incorporated into the normalisation process to further compensate for the spectral variability amongst each word. Finally, the present study employed behavioral measures to quantify performance. That is, participants had to give a voluntary verbal response. It is widely acknowledged that behavioral paradigms such as the one employed in this study, are not sensitive measures of auditory processing alone because they tap into the listeners higher level cognitive processing of attention, memory, and intelligence (Dawes & Bishop, 2009).

Our novel method of normalisation provides advantages over other, more common, methods of normalisation as it takes into consideration both the SRT and the slope of the test word, relative to the average performance. In more common normalisation methods, the SRT is adjusted by some specified value. This does not account for differences in the slope of the function, and while the SRTs of each word may be the same, fixed changes in stimulus intensity will result in different levels of word recognition performance dependent on the individual psychometric function of each test item. This is particularly an issue in adaptive speech tests, such as speech-in-noise tests, in which the level of the stimulus is altered depending on the individual's previous response. If the speech items in an adaptive test have different psychometric slope functions, there will be differences in word recognition performance depending on stimulus intensity and the test word. Therefore, the reliability and validity of a speech test is compromised when the psychometric slopes are dissimilar, as each speech item is not directly comparable when stimulus intensity is altered. Our new method of normalisation has the potential to solve this issue by adjusting the slope and the SRT of each test word to some specified standard, which in this study, was the average word recognition performance of all 50 words in the UCAST- FW. We choose to normalise all words in the UCAST-FW against the average performance. The rational behind this was that the average performance took into consideration the word recognition performance of all 50 words in the UCAST-FW list. That is, that the differences in word recognition performance due to spectral variability or lexical contribution were compiled, giving an average representation of the combined performance of the entire word list. This method was preferred over the selection of some other value as it provided an objective value to which we could normalise each word against. The generation of some other subjective value (a value or function that we considered to be better than the average) could have resulted in some words undergoing large amounts of adjustments in the level of filtering to create equal performance. Large changes in the level of filtering can drastically alter the intelligibility of words. Therefore, by taking the average performance, we can assume that the most difficult and the most easily recognizable words will undergo similar levels of adjusting, rather than certain words having large adjustments. The idea was that by normalising against the average performance, that the extreme words received as small of an adjustment as possible, to reduce the chance of completely altering the intelligibility of the test words.

One discrepancy with our method of normalisation is that the normalisation process is different for each word and at each filter frequency, as well as the previous responses of the participant. Each adjustment needs to be made on a case-by-case basis, which requires the use of algorithms to calculate the desired outcome, and is not a fixed level change in filter frequency. This means that this normalisation process needs its own software, and cannot simply be put on a CD. However, there are many cases of speech test that require this type of formatting, such as the QuickSIN.

7 References

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom–up and top–down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance, 27*(5), 1072.

American Speech-Language-Hearing, A. (1988). Determining threshold level for speech.

American Speech-Language-Hearing, A. (2005). (Central) auditory processing disorders.

Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions *Speech processing in the auditory system* (pp. 231-308): Springer.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language, 44*(4), 568-591.

Bamiou, D. E., Musiek, F. E., & Luxon, L. M. (2001). Aetiology and clinical presentations of auditory processing disorders—a review. *Archives of disease in childhood, 85*(5), 361-365.

Bartlett, E. L., Stark, J. M., Guillery, R. W., & Smith, P. H. (2000). Comparison of the fine structure of cortical and collicular terminals in the rat medial geniculate body. *Neuroscience, 100*(4), 811-828.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C., & Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of cognitive neuroscience, 10*(4), 536-540.

Bellis, T. J. (2002). *When the brain can't hear: Unraveling the mystery of auditory processing disorder*: Simon and Schuster.

Bellis, T. J. (2004). Understanding auditory processing disorders in children. *ASHA-American Speech-Language-Hearing Association [artigo online]. Disponível em: http//www. asha. org/public/hearing/disorders/understand-apd-hild. htm (25 Abr 2006).[Links]*.

Bocca, E., Calearo, C., & Cassinari, V. (1954). A new method for testing hearing in temporal lobe tumours: preliminary report. *Acta oto-laryngologica, 44*(3), 219-221.

Boothroyd, A. (1978). Speech perception and sensorineural hearing loss. *Auditory management of the hearing impaired child*, 117-144.

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America, 111*(6), 2801-2810.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *science, 315*(5820), 1860-1862.

Cacace, A. T., & McFarland, D. J. (1998). Central auditory processing disorder in school-aged children: A critical review. *Journal of Speech, Language, and Hearing Research, 41*(2), 355-373.

Cacace, A. T., & McFarland, D. J. (2005). The importance of modality specificity in diagnosing central auditory processing disorder. *American journal of audiology, 14*(2), 112-123.

Cacace, A. T., & McFarland, D. J. (2008). *Controversies in central auditory processing disorder*: Plural Publishing.

Caspary, D. M., Ling, L., Turner, J. G., & Hughes, L. F. (2008). Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *Journal of Experimental Biology, 211*(11), 1781-1791.

Chechik, G., Anderson, M. J., Bar-Yosef, O., Young, E. D., Tishby, N., & Nelken, I. (2006). Reduction of information redundancy in the ascending auditory pathway. *Neuron, 51*(3), 359-368.
Chermak, G. D., Hall, J. W., & Musiekl, F. E. (1999). Differential Diagnosis and Management of Central Auditory Processing Disorder and. *Journal of the American Academy of Audiology, 10*(6), 289-303.
Ching, T. Y. C., Dillon, H., & Byrne, D. (1998). Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *The Journal of the Acoustical Society of America, 103*(2), 1128-1140.
Clopper, C. G., Pisoni, D. B., & Tierney, A. T. (2006). Effects of open-set and closed-set task demands on spoken word recognition. *Journal of the American Academy of Audiology, 17*(5), 331-349.
Cole, R. A., & Rudnicky, A. I. (1983). What's new in speech perception? The research and ideas of William Chandler Bagley, 1874–1946. *Psychological Review, 90*(1), 94.
Cope, T. E., Baguley, D. M., & Griffiths, T. D. (2015). The functional anatomy of central auditory processing. *Practical neurology*, practneurol-2014-001073.
Cowan, J., Rosen, S., Moore, D., Cacace, A., & McFarland, D. (2009). Putting the auditory processing back into auditory processing disorder in children. *Controversies in central auditory processing disorder*, 187-198.
Dawes, P., & Bishop, D. (2009). Auditory processing disorder in relation to developmental disorders of language, communication and attention: a review and critique. *International Journal of Language & Communication Disorders, 44*(4), 440-465.
Dawes, P., Bishop, D. V. M., Sirimanna, T., & Bamiou, D.-E. (2008). Profile and aetiology of children diagnosed with auditory processing disorder (APD). *International journal of pediatric otorhinolaryngology, 72*(4), 483-489.
Decker, T. N., & Nelson, P. W. (1981). Maturation effects on the synthetic sentence identification-ipsilateral competing message. *Ear and hearing, 2*(4), 165-169.
Dorman, M. F., Dankowski, K., McCandless, G., Parkin, J. L., & Smith, L. (1991). Vowel and consonant recognition with the aid of a multichannel cochlear implant. *The Quarterly Journal of Experimental Psychology, 43*(3), 585-601.
Elliott, E. M., Bhagat, S. P., & Lynn, S. D. (2007). Can children with (central) auditory processing disorders ignore irrelevant sounds? *Research in Developmental Disabilities, 28*(5), 506-517.
Elliott, L. L., & Katz, D. (1979). Development of northwestern university children's perception of speech (nu-chips) test. *St. Louis: Auditec*.
Esplin, J., & Wright, C. (2014). Auditory Processing Disorder: New Zealand.
Farah, R., Brown, D. K., & Keith, R. W. (2013). Developmental Central Auditory Processing Disorders *Encyclopedia of Otolaryngology, Head and Neck Surgery* (pp. 676-688): Springer.
Farrer, S. M., & Keith, R. W. (1981). Filtered word testing in the assessment of children's central auditory abilities. *Ear and hearing, 2*(6), 267-269.
Field, A. (2013). *Discovering statistics using IBM SPSS statistics*: Sage.
Fuchs, P. A., Glowatzki, E., & Moser, T. (2003). The afferent synapse of cochlear hair cells. *Current opinion in neurobiology, 13*(4), 452-458.
Gathercole, S. E., & Martin, A. J. (1996). Interactive processes in phonological memory. *Models of short-term memory*, 73-100.
Gelfand, S. A. (1998). Optimizing the reliability of speech recognition scores. *Journal of Speech, Language, and Hearing Research, 41*(5), 1088-1102.

Gelfand, S. A. (2003). Tri-word presentations with phonemic scoring for practical high-reliability speech recognition assessment. *Journal of Speech, Language, and Hearing Research, 46*(2), 405-412.
Griffiths, T. D. (2002). Central auditory processing disorders. *Current opinion in neurology, 15*(1), 31-33.
Grossberg, S. (1999). The link between brain learning, attention, and consciousness. *Consciousness and cognition, 8*(1), 1-44.
Hines, T. (1999). A demonstration of auditory top-down processing. *Behavior Research Methods, Instruments, & Computers, 31*(1), 55-56.
Hudgins, C. V., Hawkins, J. E., Kaklin, J. E., & Stevens, S. S. (1947). The development of recorded auditory tests for measuring hearing loss for speech. *The Laryngoscope, 57*(1), 57-89.
Jerger, J., & Musiek, F. (2000). Report of the consensus conference on the diagnosis of auditory processing. *Journal of the American Academy of Audiology, 11*(9), 467-474.
Jerger, J. F. (1960). Observations on auditory behavior in lesions of the central auditory pathways. *AMA archives of otolaryngology, 71*(5), 797-806.
Keith, R. W. (2000). Development and standardization of SCAN-C test for auditory processing disorders in children. *Journal of the American Academy of Audiology, 11*(8), 438-445.
Kimura, D. (1964). Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology, 16*(4), 355-358.
Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R* (Vol. 32): Springer Science & Business Media.
Levitt, H. (1971). Transformed up ‐ down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*(2B), 467-477.
Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing, 19*(1), 1.
Madell, J. R., & Flexer, C. A. (2008). *Pediatric audiology: Diagnosis, technology, and management*: Thieme New York:.
Martin, F. N., Champlin, C. A., & Perez, D. D. (2000). The question of phonetic balance in word recognition testing. *Journal-american academy of audiology, 11*(9), 509-513.
Masterton, R. B., & Imig, T. J. (1984). Neural mechanisms for sound localization. *Annual review of physiology, 46*(1), 275-287.
Meister, H. (2017). Speech audiometry, speech perception, and cognitive functions. *Hno, 65*(1), 1-4.
Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of neurophysiology, 56*(3), 640-662.
Miller, C. A., & Wagstaff, D. A. (2011). Behavioral profiles associated with auditory processing disorder and specific language impairment. *Journal of communication disorders, 44*(6), 745-763.
Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27*(2), 338-352.
Moore, D. R. (2006). Auditory processing disorder (APD): Definition, diagnosis, neural basis, and intervention. *Audiological Medicine, 4*(1), 4-11.
Murray, S. L. (2012). Development of the New Zealand Stimuli for the University of Canterbury Adaptive Speech Test-Filtered Words (UCAST-FW).

Musiek, F. E., & Baran, J. A. (1987). Central auditory assessment: thirty years of challenge and change. *Ear and hearing, 8*(4), 22S-35S.
Musiek, F. E., & Chermak, G. D. (2013). *Handbook of central auditory processing disorder, volume I: auditory neuroscience and diagnosis* (Vol. 1): Plural Publishing.
Neville, K., Foley, M., & Gertner, A. (2011). Understanding and Identifying the Child at Risk for Auditory Processing Disorders A Case Method Approach in Examining the Interdisciplinary Role of the School Nurse. *The Journal of School Nursing, 27*(1), 22-33.
Nissen, S. L., Harris, R. W., Channell, R. W., Conklin, B., Kim, M., & Wong, L. (2011). The development of psychometrically equivalent Cantonese speech audiometry materials. *International journal of audiology, 50*(3), 191-201.
O'Beirne, G. A. (2009). University of Canterbury Adaptive Speech Test (UCAST)(Version 202). *Christchurch: University of Canterbury*.
O'Beirne, G. A., McGaffin, A. J., & Rickard, N. A. (2012). Development of an adaptive low-pass filtered speech test for the identification of auditory processing disorders. *International journal of pediatric otorhinolaryngology, 76*(6), 777-782.
Palfery, T. D., & Duff, D. (2007). Central auditory processing disorders: review and case study. *Axone, 28*(3), 20-23.
Pisoni, D. B. (2000). Cognitive factors and cochlear implants: Some thoughts on perception, learning, and memory in speech perception. *Ear and hearing, 21*(1), 70.
Ramkissoon, I., Proctor, A., Lansing, C. R., & Bilger, R. C. (2002). Digit speech recognition thresholds (SRT) for non-native speakers of English. *American Journal of Audiology, 11*(1), 23-28.
Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature neuroscience, 12*(6), 718-724.
Rees, A. (2009). Medial Geniculate Body *Encyclopedia of Neuroscience* (pp. 2275-2279): Springer.
Rickard, N. A., Heidtke, U. J., & O'Beirne, G. A. (2013). Assessment of auditory processing disorder in children using an adaptive filtered speech test. *International journal of audiology, 52*(10), 687-697.
Sahli, S. (2009). Auditory processing disorder in children: definition, assessment and management. *Int. Adv. Otol, 5*(1), 104-115.
Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348-351.
Schnupp, J., Nelken, I., & King, A. (2011). *Auditory neuroscience: Making sense of sound*: MIT press.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *science, 270*(5234), 303.
Silberer, A. B., Bentler, R., & Wu, Y.-H. (2015). The importance of high-frequency audibility with and without visual cues on speech recognition for listeners with normal hearing. *International journal of audiology, 54*(11), 865-872.
Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Training to improve hearing speech in noise: biological mechanisms. *Cerebral Cortex*, bhr196.
Storkel, H. L., & Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *clinical linguistics & phonetics, 14*(6), 407-425.
Teatini, G. P. (1970). Sensitized speech tests: Results in normal subjects. *Danavox Foundation, Odense, Denmark*.

Tollin, D. J., & Yin, T. C. T. (2005). Interaural phase and level difference sensitivity in low-frequency neurons in the lateral superior olive. *Journal of Neuroscience, 25*(46), 10648-10657.
Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (2000). English speakers' sensitivity to phonotactic patterns. *Papers in laboratory phonology V: Acquisition and the lexicon*, 269-282.
Vitevich, M. S., & Luce, P. A. (1998). When words compete: levels of processing in spoken word recognition. *Psychological Science, 9*, 325-329.
Vitevitch, M. S. (2003). The influence of sublexical and lexical representations on the processing of spoken words in English. *clinical linguistics & phonetics, 17*(6), 487-499.
Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*(3), 374-408.
Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and speech, 40*(1), 47-62.
Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and language, 68*(1), 306-311.
Wagener, K., Josvassen, J. L., & Ardenkjær, R. (2003). Design, optimization and evaluation of a Danish sentence test in noise: Diseño, optimización y evaluación de la prueba Danesa de frases en ruido. *International journal of audiology, 42*(1), 10-17.
Weinberger, N. M., Ashe, J., & Edeline, J.-M. (1994). Learning-induced receptive field plasticity in the auditory cortex: specificity of information storage *The memory system of the brain* (pp. 590-635): World Scientific.
Weisleder, P., & Hodgson, W. R. (1989). Evaluation of four Spanish word-recognition-ability lists. *Ear and hearing, 10*(6), 387-393.
Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology, 18*(1), 1-10.
Whitelaw, G. M., & Yuskow, K. (2006). Neuromaturation and neuroplasticity of the central auditory system. *An introduction to auditory processing disorders in children*, 21-38.
Willeford, J. A., & Billger, J. (1978). Auditory perception in children with learning disabilities. *Handbook of clinical audiology, 2*.
Wilson, R. H., & Carter, A. S. (2001). Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials. *Journal of the American Academy of Audiology, 12*(1), 7-14.
Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing a. *The Journal of the Acoustical Society of America, 131*(2), 1465-1479.

8 Appendices

Appendix A: Ethical Approval

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Rebecca Robinson
Telephone: +64 03 369 4588, Extn 94588
Email: human-ethics@canterbury.ac.nz

Ref: 2017/02/ERHEC-LR

8 February 2017

Jasmine Gibbins
Communication Disorders
UNIVERSITY OF CANTERBURY

Dear Jasmine

Thank you for submitting your low risk application to the Educational Research Human Ethics Committee for your research proposal titled "Normative Data for the University of Canterbury Adaptive Speech Test-Filtered Word (UCAST-FW) of Auditory Processing Disorders (APD) in Children".

I am pleased to advise that this application has been reviewed and I confirm support of the School's approval for this project.

With best wishes for your project.

Yours sincerely

pp R. Robinson

Dr Patrick Shepherd
Chair
Educational Research Human Ethics Committee

Please note that ethical approval relates only to the ethical elements of the relationship between the researcher, research participants and other stakeholders. The granting of approval by the Educational Research Human Ethics Committee should not be interpreted as comment on the methodology, legality, value or any other matters relating to this research.

F E S

Appendix B: Recruitment and Consent

B1: Study advertisment for participant recruitment

RESEARCH PARTICIPANTS NEEDED

The Development of a Psychometrically Equivalent Word list for the Normalization of the University of Canterbury Adaptive Speech Test- Filtered Word (UCAST-FW)

PROJECT INFO

We are seeking participants to get involved in a study to help in the development of a speech test for the diagnosis of auditory processing disorder (APD). We need normal hearing adults to perform 2 simple listening tasks at the University of Canterbury Speech and Language Clinic, in the Communication Disorders Department.

1 ARE YOU...

BETWEEN 18-40 YEARS OF AGE

HAVE NORMAL HEARING

HAVE NO KNOWN LANGUAGE OR LEARNING DIFFICULTIES

A NATIVE SPEAKER OF NZ ENGLISH

2 WHAT I NEED:

45-60 MINUTES OF YOUR TIME

3 BENIFITS FOR YOU:

A FREE HEARING TEST

$10 VOUCHER

PLENTY OF SNACKS

B2: Information sheet given to each participant prior to testing (page 1 of 3)

University of Canterbury
Department of Communication Disorders

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Project Name: The development of a psychometrically equivalent word list for the normalization of the University of Canterbury Adaptive Speech Test- FW (UCAST-FW) for diagnosis of auditory processing disorder.

AIM OF THE PROJECT

You are invited to take part in this research project to assist in the development of a speech test used in the diagnosis of auditory processing disorder (APD). APD is a term used to describe individuals with normal hearing, who have auditory-based receptive communication or language learning problem. APD affects "what we do with what we hear".

This project is evaluating the reliability and validity of the words in this UCAST-FW. The UCAST-FW test is a speech test where the words in the test are filtered to make them more difficult to hear. This test will examine how well participants recognize words in the test under different levels of filtering.

To carry out this research, we are looking for 30 people between the ages of 18-40 years, who are native New Zealand speakers of English, and have no known speech, language or learning difficulties.

If you choose to undertake this project you will be asked to complete the following tasks in our research laboratory:

1) **A standard hearing test:** Your ear canals will be examined to determine ear health, and a hearing test will be conducted via pure tone audiometry; a standard practice for testing hearing, where you will be asked to respond to a beep (Approx. 10 minutes). If any hearing difficulties are identified, a full hearing assessment can be organized at the Speech and Hearing Clinic, Department of Communication Disorders, University of Canterbury.

B2: Information sheet given to each participant prior to testing (page 2 of 3)

2) **Listening experiment 1**: You will be placed in front of a computer with a set of headphones. You will hear a word presented via the headphones, with four pictures on the computer screen. You must select the picture that corresponds to the word you hear, or take your best guess. This process will repeat with the words sometimes sounding more muffled, and sometimes more clear (approx. 25 minutes).

3) **Listening experiment 2.** Similar to the experiment 1, you will hear a word presented through the headphones, however, this time you will repeat the word you heard back to me. Again this will repeat, with words being either more or less difficult to hear (approx. 30 minutes).

WITHDRAWL & CONFIDENTIALITY

You have the right to withdrawal from the project at any time, including any information provided.

The results of this project may be published, but you will have complete confidentiality of data gathered during this research. Your identity will not be made public with out direct consent. To ensure anonymity and confidentiality, the information gathered will be assigned a code number and all identifiable information will be removed. Data will be kept in a locked filling cabinet within a lockable room in the Department of Communication Disorders at the University of Canterbury.

The project being carried out is a requirement for the completion of a Master of Audiology Degree by Jasmine Gibbins, under the supervision of Doctor Greg O'Beirne, who can be contacted at the University of Canterbury on +64 3 364 2987 ext. 7085. We will be pleased to discuss any concerns you may have about the participation in the project.

B2: Information sheet given to each participant prior to testing (page 3 of 3)

<u>Reimbursement for participation</u>

On completion of the assessment you will receive a $10 MTA voucher for reimbursement for your time.

<u>Ethics</u>

The University of Canterbury Human Ethics Committee has approved this project.

Jasmine Gibbins
Master of Audiology Student
Department of Communication
Disorders
University of Canterbury
Private Bag, 4800
Christchurch, New Zealand
Email: Jgi52@uclive.ac.nz

B3: Consent form signed by all participants

UNIVERSITY OF CANTERBURY
DEPARTMENT OF COMMUNICATION DISORDERS

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

The Development of a Psychometrically Equivalent Word List for the Normalization of the University of Canterbury Adaptive Speech Test- FW (UCAST-FW)

I have read and understood the information given to me about the research project named above. I have had a chance to ask questions and have had them answered. I understand that my participation in this project is voluntary and that I am free to withdraw at any time without giving a reason. I understand that the results of the session will be written down. I understand that the information you collect from me will remain confidential and will be securely stored at the university. I understand that any presentations or publications resulting from this project will not refer to me by name. I understand that this project has been reviewed ***and approved*** by the University of Canterbury Human Ethics Committee. On this basis, I agree to participate in this research project.

My Name (please print):...

My Signature: ....

Date...

Main Researcher: Jasmine Gibbins Email: Jgi52@uclive.ac.nz
University of Canterbury Private Bag 4800, Christchurch 8140, New Zealand. www.canterbury.ac.nz

Appendix C: Participant Instructions

UCAST closed instructions provided to each participant

"*Shorty, I will place these headphones over your ears. Through the headphones, you will hear a man's voice. He will say a word, like cat. The word will sound a bit muffled like this."*

(The tester who uses their hand over their mouth to simulate the filtered word provided a verbal example of the word to the participant.)

*"After you hear the word, four pictures will appear on the computer screen in front of you. Your task is to match the word you heard through the headphones with one of the pictures shown on the screen. * Uses the mouse to click on the screen- like this."*

(A demonstration of how to use the mouse to select a word on the screen was provided for the participant by the tester.)

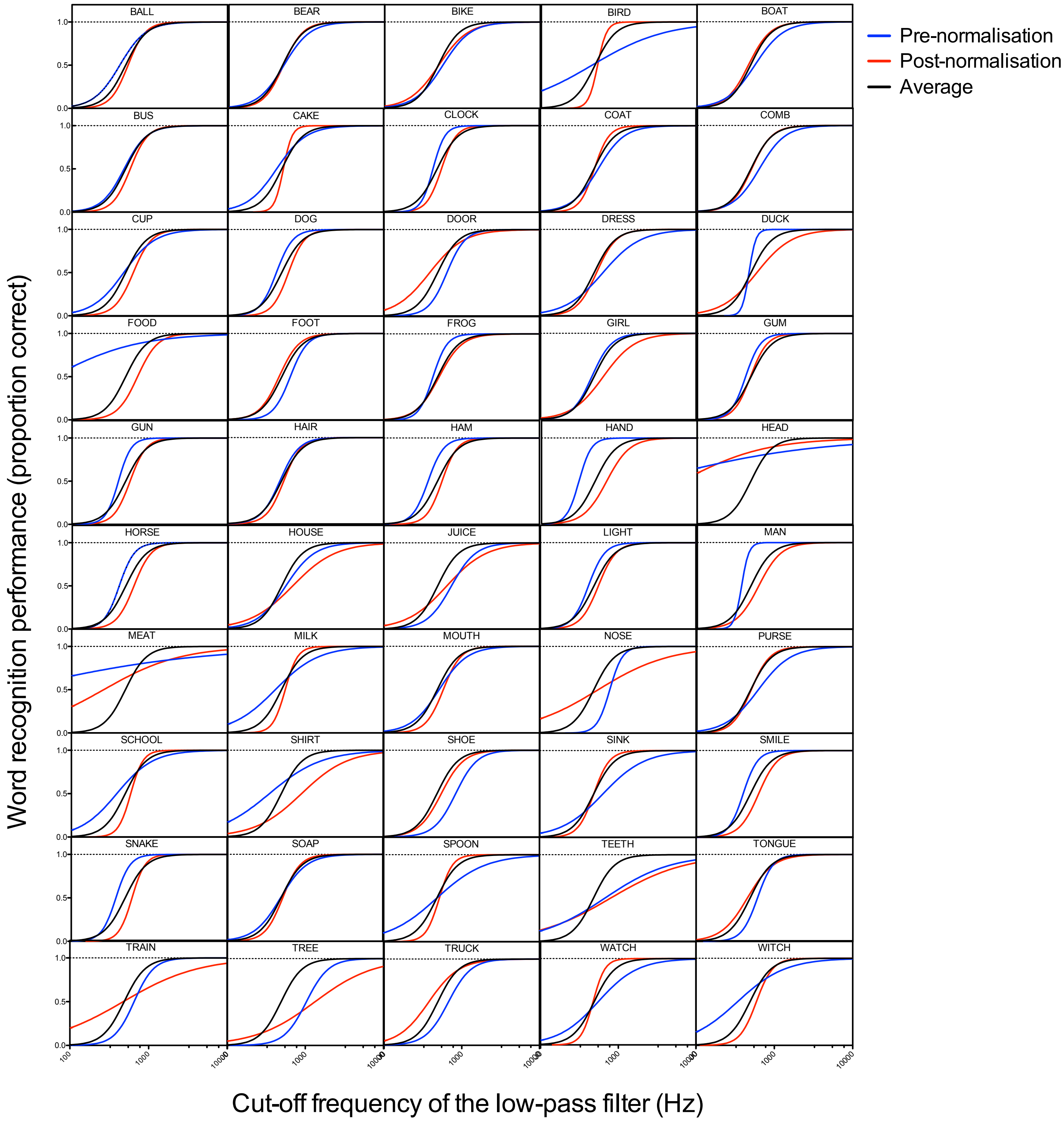
"After you choose a picture, you will hear another word through the headphones. Again, you choose the picture that matches the word. Sometime the word will be much muffled and might be difficult to tell what is being said. The word cannot be repeated, so if you are not sure as to what was said, just take a guess. The test will take about 25 minutes and will stop by itself. Do you have any questions?"

UCAST open set instructions provided to each participant

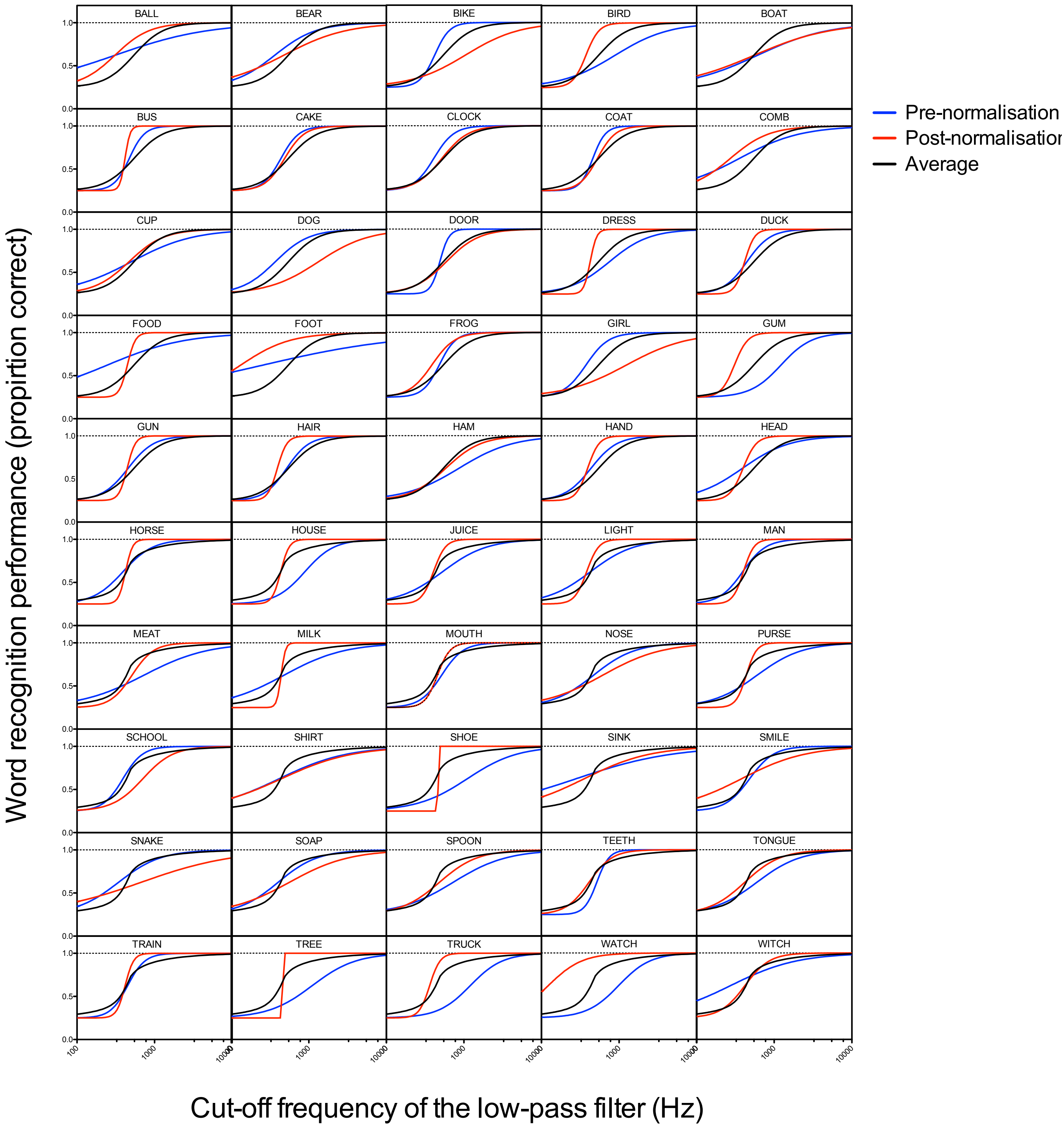
"*Again, I will place these headphones over your ears. Just as before, you will hear a man's voice through the headphones, saying a single syllable word. This time I want you to repeat the word that you heard back to me. The words will be muffled. If you are unsure as to what you heard, I want you to take your best guess. Even if what you repeat, is a nonsense syllable. This test will take about 40 minutes. We will take a break at half way so you don't get too tired"*

Appendix D: Additional research data

D1: Psycometric curves for all words in the UCAST-FW test for pre- and post-normalisation conditions of the open set paradigm.



D2: Psychometric curves for all words in the UCAST-FW test for pre- and post-normalisation conditions of the closed set paradigm.



D3: Percent deviation from the average performance for SRT and slope values for open set

Average SRT = 489.4 Hz, average slope = 50.22 5/octave. 0% change indicates no deviation from the average negative values indicate SRT values that are greater than the average.

	Pre-normalisation		Post-normalisation		Pre-normalisation		Post-normalisation	
Word	**SRT**	**% change from average**	**SRT**	**% change from average**	**Slope**	**% change from average**	**Slope**	**% change from average**
Ball	411.6	18.89	525.5	-6.87	50	0.38	54.1	-7.15
Bear	509.5	-3.95	500.8	-2.28	50	0.50	52.9	-5.10
Bike	560.7	-12.72	497.1	-1.55	49.4	1.64	48.8	2.93
Bird	449.3	8.92	529.2	-7.53	39.6	26.69	60.6	-17.17
Boat	556.8	-12.11	459.4	6.53	49.1	2.39	51.8	-3.09
Bus	467.8	4.61	561.4	-12.82	51.5	-2.45	54.0	-7.03
Cake	420.1	16.49	512.5	-4.51	48.8	2.93	61.0	-17.63
Clock	433	13.03	556.1	-12.00	57.9	-13.22	55.8	-9.92
Coat	553.9	-11.64	494.9	-1.12	50	0.44	55.8	-9.97
Comb	627.8	-22.05	499.2	-1.97	50.1	0.24	52.6	-4.47
Cup	462.6	5.80	613.7	-20.25	48.1	4.39	53.5	-6.08
Dog	418.4	16.96	598.9	-18.28	54.6	-7.99	54.9	-8.46
Door	644.3	-24.05	386	26.78	53.6	-6.22	47.2	6.38
Dress	635.6	-23.00	525.5	-6.87	46.3	8.42	52.8	-4.94
Duck	461.3	6.10	581.9	-15.89	63.5	-20.95	47.1	6.69
Food	56.7	762.79	695.9	-29.67	38.3	31.05	52.6	-4.54
Foot	629.6	-22.27	452	8.28	54.2	-7.38	52.7	-4.67
Frog	427.5	14.47	511.4	-4.31	56.1	-10.51	51.2	-1.97
Girl	458.1	6.83	643.6	-23.96	52.9	-4.98	48.1	4.47
Gum	417.2	17.30	481.8	1.58	55.6	-9.68	54.5	-7.79
Gun	395.4	23.77	557.9	-12.27	57.2	-12.26	54.8	-8.42
Hair	466.2	4.98	522.3	-6.31	53.1	-5.44	54.2	-7.31
Ham	370.4	32.14	559.8	-12.57	55.1	-8.87	55.4	-9.35
Hand	303.4	61.28	689.3	-29.00	57.5	-12.65	51.2	-1.95
Head	22.4	2088.63	63.2	674.09	31.6	58.92	38.4	30.92
Horse	403	21.43	623.2	-21.48	55.4	-9.42	54.1	-7.09
House	551.7	-11.29	676.8	-27.69	49	2.51	45.0	11.53
Juice	737	-33.60	640.7	-23.62	50.7	-0.97	45.8	9.72
Light	415.8	17.70	558.1	-12.31	54.7	-8.21	53.9	-6.84
Man	371.1	31.88	617	-20.68	62.8	-20.01	51.5	-2.56
Meat	15.9	2972.05	257.3	90.17	30.3	65.85	39.2	27.98
Milk	407.6	20.05	546.1	-10.38	45	11.62	57.5	-12.65
Mouth	517.3	-5.39	584.8	-16.31	49.3	1.80	54.0	-7.02
Nose	784.9	-37.65	554.8	-11.79	57.3	-12.29	40.0	25.61
Purse	610.9	-19.89	496.4	-1.40	47.8	5.08	53.3	-5.69
School	381	28.46	575.4	-14.95	46.7	7.61	57.5	-12.72
Shirt	344.5	42.06	889.2	-44.96	43.1	16.63	44.4	13.24
Shoe	848.8	-42.34	555.2	-11.86	52	-3.39	51.5	-2.50
Sink	641.3	-23.69	490	-0.13	45.3	10.79	55.2	-8.94
Smile	394.9	23.92	622.4	-21.37	55.9	-10.21	53.2	-5.57
Snake	372.4	31.43	598.7	-18.26	56.1	-10.51	57.0	-11.82
Soap	482.6	1.40	514	-4.78	50	0.50	54.1	-7.10
Spoon	488.8	0.13	527.5	-7.22	43.9	14.45	56.9	-11.72
Teeth	683.3	-28.38	807.8	-39.42	40.9	22.91	39.6	26.91
Tongue	607.8	-19.48	451.9	8.29	55.2	-9.02	50.1	0.30
Train	654.6	-25.24	476.1	2.80	52.9	-5.05	39.5	27.17
Tree	1031.3	-52.55	1320.1	-62.93	51.7	-2.86	41.9	19.89
Truck	672.6	-27.24	374.7	30.61	51.4	-2.30	48.5	3.61
Watch	551.8	-11.31	474.3	3.19	45.8	9.60	57.9	-13.19
Witch	338.56	44.55	590.82	-17.17	81.57	124.32	238.31	-23.22
Average	489.37	0.00	489.37	0.00	182.97	0.00	182.97	0.00

D4: Percent deviation from the average performance for SRT and slope values for closed set

Average SRT = 547.3, average slope = 47.1. 0% change indicates no deviation from the average negative values indicate SRT values that are greater than the average.

	Pre-normalisation		Post-normalisation			Pre-normalisation		Post-normalisation	
Word	**SRT**	**% Change from average**	**SRT**	**% Change from average**		**Slope**	**% Change from average**	**Slope**	**% Change from average**
Ball	316.8	72.78	311.6	75.64		37.2	26.38	47.2	-0.30
Bear	366.1	49.50	478.8	14.31		45.4	3.68	41.2	14.09
Bike	421.5	29.83	1059.0	-48.32		55.6	-15.44	42.7	10.14
Bird	892.9	-38.70	375.5	45.74		42.8	9.83	56.9	-17.24
Boat	609.5	-10.21	553.1	-1.05		40.1	17.45	39.3	19.78
Bus	474.4	15.37	417.4	31.13		55.4	-15.01	66.5	-29.19
Cake	437.9	24.98	488.8	11.97		53.6	-12.27	52.8	-10.82
Clock	411.6	32.99	524.6	4.33		52.7	-10.67	49.9	-5.79
Coat	460.5	18.84	528.9	3.49		57.2	-17.73	53.7	-12.45
Comb	357.2	53.22	272.7	100.73		41.5	13.48	46.0	2.28
Cup	507.8	7.77	471.6	16.05		41.4	13.78	47.1	-0.06
Dog	370.4	47.76	1335.8	-59.03		47.5	-0.88	43.5	8.19
Door	490.7	11.54	599.1	-8.65		60.5	-22.27	47.7	-1.40
Dress	755.2	-27.53	426.0	28.47		45.8	2.84	64.1	-26.59
Duck	445.4	22.87	416.1	31.54		51.2	-8.03	59.8	-21.26
Food	249.1	119.68	447.6	22.27		39.1	20.49	61.8	-23.92
Foot	258.1	112.08	132.9	311.94		33.2	41.59	43.7	7.72
Frog	483.0	13.31	388.5	40.89		54.7	-13.97	51.4	-8.39
Girl	375.9	45.59	1251.6	-56.27		51.5	-8.64	41.5	13.26
Gum	1224.5	-55.30	307.2	78.19		48.5	-2.91	56.8	-17.14
Gun	459.1	19.21	445.5	22.84		50.2	-6.26	61.6	-23.64
Hair	518.8	5.49	393.8	39.00		51.5	-8.68	60.1	-21.77
Ham	872.3	-37.26	587.9	-6.90		42.7	10.08	47.0	0.13
Hand	436.0	25.53	394.8	38.63		50.0	-5.92	58.3	-19.27
Head	390.0	40.32	396.8	37.94		44.1	6.69	56.8	-17.12
Horse	392.3	39.51	422.6	29.50		48.6	-3.17	63.3	-25.67
House	893.0	-38.71	435.7	25.62		48.9	-3.74	61.3	-23.26
Juice	534.1	2.48	406.2	34.74		44.7	5.19	57.8	-18.63
Light	432.5	26.55	384.1	42.47		44.8	4.95	57.8	-18.64
Man	393.5	39.07	401.4	36.35		51.2	-8.18	59.1	-20.42
Meat	729.9	-25.01	512.1	6.87		41.1	14.42	51.8	-9.20
Milk	479.4	14.16	440.1	24.35		41.6	13.16	68.2	-30.96
Mouth	517.7	5.71	456.9	19.78		52.0	-9.45	57.3	-17.93
Nose	472.1	15.92	620.5	-11.79		45.4	3.68	41.8	12.45
Purse	604.7	-9.50	436.3	25.44		44.5	5.64	59.4	-20.80
School	391.3	39.85	707.3	-22.62		52.8	-10.89	48.5	-3.03
Shirt	416.4	31.43	443.2	23.49		40.4	16.43	39.9	17.83
Shoe	1111.1	-50.74	463.9	17.97		43.5	8.11	90.4	-47.97
Sink	282.8	93.51	359.2	52.37		36.9	27.51	40.7	15.63
Smile	483.3	13.25	381.7	43.40		50.7	-7.27	41.0	14.64
Snake	344.4	58.91	675.1	-18.93		45.2	4.21	37.4	25.97
Soap	404.1	35.45	565.2	-3.16		45.8	2.66	41.7	12.75
Spoon	731.6	-25.19	486.1	12.59		42.8	9.85	46.2	1.86
Teeth	518.7	5.51	412.8	-39.42		57.5	-18.17	51.2	-8.03
Tongue	601.0	-8.93	419.4	8.29		44.8	4.93	46.8	0.45
Train	450.0	21.61	417.3	2.80		54.1	-13.08	61.3	-23.27
Tree	1111.2	-50.75	460.7	-62.93		44.9	4.84	89.9	-47.66
Truck	1212.4	-54.86	362.5	30.61		47.9	-1.82	59.2	-20.50
Watch	992.7	-44.87	123.7	3.19		48.1	-2.18	47.4	-0.78
Witch	278.5	96.50	423.2	-17.17		40.7	15.72	50.6	-7.07
Average	547.3	16.61	484.5	32.98		37.04	1.62	489.37	-8.46